(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/066821 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/03551

(22) International Filing Date: 4 February 2003 (04.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/354,178	4 February 2002 (04.02.2002)	US
60/384,302	30 May 2002 (30.05.2002)	US
60/388,325	13 June 2002 (13.06.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/354,178 (CIP)
Filed on	4 February 2002 (04.02.2002)
US	60/384,302 (CIP)
Filed on	30 May 2002 (30.05.2002)
US	60/388,325 (CIP)
Filed on	13 June 2002 (13.06.2002)

(71) Applicant *(for all designated States except US)*: **THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE** [US/US]; 111 Market Place, Suite 906, Baltimore, MD 21202 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KOLODKIN, Alex, L.** [US/US]; 5728 Cross Country Blvd., Baltiore, MD (US). **TERMAN, Jon, Richard** [US/US]; 20 Red Hearth Ct., Baltimore, MD 21227 (US). **MAO, Tianyi** [CN/US]; 9228 Kingstree Rd., #304, Parkville, MD 21234 (US). **PASTERKAMP, Ronald, Jeroen** [NL/US]; 681A St. Paul Street, baltimore, MD 21202 (US). **YU, Hung-Hsiang** [—/US]; 15923 Hwy 99, A105, Lynnwood, WA 98037 (US).

(74) Agent: **HAILE, Lisa, A.**; Gary Cary Ware & Friedenrich LLP, 4365 Executive Drive, Suite 1100, San Diego, CA 92121-2133 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/066821 A2

(54) Title: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPTIDES, AND METHODS OF USING THE SAME

(57) Abstract: The present invention provides MICAL and MICAL-Like polypeptides and polynucleotides. Also provided are methods that for identifying agents that affect axon growth and placement. Furthermore, provided herein are methods for affecting axon growth and placement.

MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPTIDES, AND METHODS OF USING THE SAME

STATEMENT OF GOVERNMENT SUPPORT

[0001] This invention was made in part with government support under Grant Nos. NRSA-NS11055 and NS15165 awarded by the National Institutes of Health. The United States government may have certain rights in this invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0002] This invention relates generally to polynucleotides encoding a family of polypeptides, and more specifically to polynucleotides encoding polypeptides having oxygenase activity and methods of use thereof.

BACKGROUND INFORMATION

[0003] A great deal of research has focused on identifying factors that underlie human pathological conditions and disease states. This research has focused on molecules that are important for normal human development to identify those factors whose function is compromised in human disorders. For example, during development, neurons form connections with one another and with other targets by extending processes called axons.

[0004] In order to make developmental connections, axons navigate over long distances selecting their correct pathway, finding their appropriate target area, and establishing the proper connections with their target. The means by which axons accomplish this, remains largely unknown. It is clear however, that identifying the molecular signals that enable axons to form these connections is important for developing treatments for many neurological disorders including spinal cord injury.

[0005] Many of the same molecules implicated as functioning to guide the growing axon have been shown to function in cell adhesion, cell proliferation, cytoskeletal integrity, and other aspects of normal cell migration both in the nervous system and outside it. Characterization of these molecules will identify strategies for treatment of cell migration disorders including tumorigenesis. Therefore, there remains a need to identify molecules that function to guide growing axons.

[0006] Following spinal cord injury in humans, axons fail to reestablish their connections, which results in paralysis and loss of sensation of the affected area. The factors that inhibit axons from reestablishing their connections are not known. It is interesting, however, that during development, inhibition of axon growth plays a role in forming the nervous system. Axons are guided to their targets by molecules that attract them as well as by those that inhibit (i.e., repel) them. These molecules help channel axons into appropriate areas, as well as prevent them from entering unwanted regions. However, these molecules remain largely unidentified. Therefore, there remains a need to identify molecules that inhibit or repel axon growth.

SUMMARY OF THE INVENTION

[0007] The present invention relates to a family of proteins, called MICALs, that are large, multidomain proteins expressed in axons, that interact with the neuronal plexin A receptor and are required for semaphorin 1a-PlexA-mediated repulsive axon guidance. In addition to containing several domains known to interact with cytoskeletal components, MICALs have a flavoprotein monooxygenase domain, the integrity of which is required for Sema-1a-PlexA repulsive axon guidance. The presence of these domains suggest a previously unknown role for oxidoreductases in repulsive neuronal guidance.

[0008] In one embodiment, the present invention provides an isolated polypeptide that includes a plexin interacting region and one or more of an N-terminal MICAL domain, a calponin homology domain, a LIM domain, and a proline rich region, wherein the polypeptide has monooxygenase activity, plexin interacting activity, and/or axon guidance regulatory activity. The polypeptide can also include a first variable MICAL region and a second variable MICAL region, which can form part of the proline rich region. The

polypeptide can include, for example, from N-terminal to C-terminal, an N-terminal MICAL domain, a calponin homology doxmain, a first variable MICAL region, a LIM domain, a proline rich region, and a plexin interacting region.

[0009] In one aspect, the polypeptide is a mammalian polypeptide. For example, the isolated polypeptide can be human MICAL-1, human MICAL-2, or human MICAL-3. Accordingly, the polypeptide can include an amino acid sequence as set forth in SEQ ID NO:2 (human MICAL-1), SEQ ID NO:4 (human MICAL-2), or SEQ ID NO:6 (human MICAL 3).

[0010] In another aspect, the polypeptide is a *Drosophila* MICAL polypeptide. For example, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:8 (*Drosophila* long variant) SEQ ID NO:10 (*Drosophila* medium variant), or SEQ ID NO:12 (*Drosophila* short variant).

[0011] In another embodiment, the present invention provides a MICAL-Like polypeptide. Accordingly, the isolated polypeptide includes a plexin interacting region and alternatively one or more of a calponin homology domain, a LIM domain, and a proline rich region, and wherein the polypeptide interacts with a plexin. The polypeptide can also include a first variable region and a second variable region, which can form part of the proline rich region. For example, the polypeptide includes, from N-terminal to C-terminal, a calponin homology domain, a first variable MICAL region, a LIM domain, a proline rich region, and a plexin interacting region.

[0012] In another aspect, the isolated polypeptide is a *Drosophila* MICAL-Like polypeptide. For example, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:18 (*Drosophila* MICAL-Like polypeptide).

[0013] In another embodiment, the present invention provides an isolated polynucleotide encoding a MICAL polypeptide or a MICAL-Like polypeptide of the present invention. The polynucleotide can encode a mammalian MICAL polypeptide, for example, human MICAL-1, human MICAL-2, human MICAL-3, human MICAL-Like 1, or human MICAL-Like 2. Accordingly, in one aspect, the polynucleotide is the coding sequence portion of

SEQ ID NO:1 (human MICAL-1 cDNA), the coding sequence portion of SEQ ID NO:3 (human MICAL-2 cDNA), the coding sequence portion of SEQ ID NO:5 (human MICAL 3 cDNA), the coding sequence portion of SEQ ID NO:13 (human MICAL-Like 1 cDNA), or the coding sequence portion of SEQ ID NO:15 (human MICAL-Like 2 cDNA).

[0014] The present invention also provides an isolated polynucleotide that selectively hybridizes to a polynucleotide encoding a MICAL polypeptide or a MICAL-Like polypeptide.

[0015] In yet another embodiment, the present invention provides a method for identifying an agent that affects axonal guidance regulatory activity. The method includes contacting a polypeptide of the present invention that has axonal guidance regulatory activity, or a cell expressing the polypeptide, for example recombinantly expressing the polypeptide, with a candidate agent. Next, axonal guidance regulatory activity or expression of the polypeptide is compared in the presence versus absence of the agent. A difference in activity or expression is indicative of an agent that affects axonal guidance regulatory activity.

[0016] In another embodiment, the present invention provides a method for affecting axonal guidance regulatory activity. The method includes contacting a cell, for example, a neuron, that expresses a polypeptide of the invention such as a MICAL polypeptide, with an agent that alters MICAL activity and, thereby, affects axonal guidance regulatory activity. In one aspect, the method is performed *in vivo* and includes inhibiting axonal guidance regulatory activity by contacting the cell with an antioxidant that inhibits MICAL activity. The axonal guidance activity is a semaphorin-mediated axonal repulsion. As such, in another embodiment, the present invention provides a method for affecting a semaphorin-mediated process by contacting a cell that expresses a MICAL polypeptide of the invention with an effective amount of an agent that modulates MICAL activity and, thereby, affects axonal guidance regulatory activity. An agent is, for example, a small molecule, a polypeptide or fragment thereof, a peptidomimetic, or an antisense polynucleotide.

[0017] In another embodiment, the present invention provides a method for treating a neurological condition in a subject, that includes contacting in the subject, a cell of the

central nervous system or the peripheral nervous system, having a disrupted axonal connection or a cell that affects axonal growth of the central nervous system or peripheral nervous system cell, with an amount of an agent that modulates the activity or expression of a MICAL polypeptide, the amount being effective to modulate axon regulatory activity, monooxygenase activity, and/or plexin interacting activity. In one aspect, the neurological condition is a spinal cord injury.

[0018] The present invention identifies exemplary flavonoids as agents that are used in methods of various embodiments of the present invention to inhibit axonal guidance regulatory activity. A variety of flavonoid anti-oxidants are known and are candidate inhibitors MICAL activity and, thereby, of axonal guidance regulatory activity such as semaphorin-mediated axonal repulsion. In one aspect of the invention, the flavonoids ECGC and EC and related gallic acid derivatives are inhibitors of semaphorin-mediated axonal repulsion.

[0019] In another aspect, the present invention provides a method for inducing regrowth of an injured process of a neuron, that includes altering the levels of reactive oxygen species or other oxidation products in the milieu of the neuron.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] Figure 1 provides a molecular characterization of *Drosophila* MICAL, MICAL expression in *Drosophila* embryonic motor axons, and co-immunoprecipitation of MICAL with neuronal PlexA.

[0021] Figure 1A provides a schematic diagram of the MICAL locus. Variable exons are indicated with asterisks and produce: 1) a "long" isoform (4723 aa); 2) a "medium" isoform (3002 aa) (spliced out exon is shown with "X" though lines); and 3) a "short" isoform (2734 aa) (spliced out exons shown by thick angled exon connector lines). The regions corresponding to clones 23 and 151 are shown.

[0022] Figure 1B provides the domain organization of the Drosophila MICAL gene. MICAL is characterized by flavin adenine dinucleotide (FAD) consensus binding motifs

(GXGXXG, DG, and GD motifs), a calponin homology domain, a LIM domain, a Proline rich region, and a coiled-coil motif.

[0023] Figure 2 provides schematic representations illustrating that the MICALs are a family of neuronally-expressed plexin-interacting proteins conserved from flies to mammals.

[0024] Figure 2A is a schematic representation of the organization of the MICAL family of proteins. Amino acid identities are indicated among vertebrate MICALs and *Drosophila* MICAL (%s within domains) and between vertebrate MICALs (%s in arrows). The black regions indicate sequence that is not well-conserved among family members and variable in length (//). Regions encoded by an ORF situated in close proximity in the genome (~10kb) but for which cDNA sequence connecting them has not yet been identified are indicated (dots).

[0025] Figure 2B provides a schematic representation of the domain organization of the MICAL-Like proteins. MICAL-like proteins have a similar domain organization as the MICALs but lack the N terminal ~500 amino acid domain. Domain alignment and amino acid identity between *Drosophila* MICAL and MICAL-like proteins is indicated (within domains) and between MICAL-like proteins (within arrows). Available D-MICAL-L cDNA and genomic DNA sequence information suggests that the D-MICAL-L protein begins just N-terminal to the CH domain. Human MICAL-L1 and MICAL-L2 are similar in overall domain organization to D-MICAL-L and do not contain the highly conserved ~500 amino acid MICAL N-terminal domain (dots indicate where molecular analysis is required to conclusively define the structural features of mammalian MICAL-L proteins; P, proline rich region; cc, coiled-coil).

[0026] Figure 3 illustrates that the MICALs contain flavoprotein monooxygenase domains required for MICAL function in *Drosophila*.

[0027] Figure 3A provides a schematic representation of three sequence motifs that define MICALs as flavoprotein monooxygenases. An alignment of MICALs with members of the flavoprotein monooxygenase family is shown in which (+) indicates that MICALs

match the consensus, (*) indicates that MICALs match the highly important conserved residues, and (.) indicates the conserved spacing of these residues within these motifs. Shading of sequences is based on ClustalX: conserved hydrophobic residues, cysteine residues, acids, and bases are shaded in dark gray; conserved proline and glycine residues are indicated with light gray shading. MICALs contain a 100% match with the consensus ADP binding region of FAD binding proteins (FAD Fingerprint 1), a well-conserved GD sequence (FAD Fingerprint 2), and a well-conserved DG motif: distinguishing features of flavoprotein monooxygenases. The proline ((*)) in the FAD fingerprint 2 is also likely to be conserved. In the upper consensus line, uppercase indicates an amino acid; h is a hydrophobic residue, s is a small residue (i.e., compact, zero, or few side chains); c is a charged residue, and x is any residue.

[0028] Figure 3B is a spectral analysis of MICAL that illustrates that MICAL is an FAD binding protein. A bacterial fusion protein consisting of the *Drosophila* MICAL flavoprotein monooxygenase (FM) domain has an absorption peak at 452 nm and a shoulder at ~358 nm (dashed line), consistent with that of 50μM free FAD (solid line) and similar in shape to spectra from other flavoproteins.

[0029] Figure 4 illustrates that flavoprotein monooxygenase inhibitors attenuate vertebrate semaphorin axonal repulsion. Inhibitors of flavoprotein monooxygenases (EGCG and EC) as well as specific inhibitors of other oxidation/reduction enzymes including nitric oxide synthase (L-NAME), xanthine oxidase (allopurinol; Allo), and mitochondrial electron transport (NADH dehydrogenase; rotenone; Rote) were tested for their ability to inhibit semaphorin-dependent repulsive axon guidance in vertebrates.

[0030] Figure 4A provides a schematic diagram of the rat D E14/15 Rat DRG explants were co-cultured with 293 cells expressing Sema 3A and grown for 48 hours. in the presence of an inhibitor or vehicle. Axonal outgrowth was determined by measuring proximal (P) and distal (D) axon lengths.

[0031] Figure 4B provides a graph illustrating that redox inhibitors do not have adverse effects on expression of Sema 3A or its biological activity. Media was collected from untransfected 293 cells (No Sema3A) or cells transfected with AP-Sema3A and grown in

the presence of vehicle (Sema3A), 25 μm EGCG (3A/EGCG), 500 μm EC (3A/EC), 500 μm L-NAME (3A/L-NAME), 500 μm Allo (3A/Allo), or 0.1 μm Rote (3A/Rote) and ligand concentration (AP activity) was determined. The media was then diluted to 1nM (to remove the active concentration of the inhibitor) and its biological activity was assayed in a growth cone collapse assay (% Collapse; n > 60 growth cones per condition). The AP activity and percentage of growth cones collapsed were similar in the presence of all compounds.

[0032] Figure 4C provides a graph that quantitates the effects of oxidation/reduction enzyme inhibitors on Sema 3A repulsion scored as the ratio of the axon lengths on the proximal and distal sides of the explant (P/D ratio), and on Sema3A-mediated growth cone collapse indicated as % collapsed growth cones (gray). In repulsion assays, outgrowth of DRG axons on the side distal to the 293 cells appeared normal. Attenuation of Sema 3A-mediated axonal repulsion was observed with the flavoprotein monooxygenase inhibitors (EGCG, and EC) in a dose-dependent manner but not with specific inhibitors of other oxidation/reduction enzymes. n's= number of DRG explants (repulsion assays) or number of growth cones scored (collapse assays; distributed over 4 different explants/condition). For Rote, n=4, however only 4 out of 12 explants survived. (**=p<0.0001; *=p<.001; paired t-test). Scale bar = 550 mm.

[0033] Figure 5 shows results of a yeast interaction assay and identification of *Drosophila* MICAL. Figure 5A provides a diagram of the Plexin A polypeptide. Figure 5B illustrates that clones 23 and 151 of the yeast interaction assay encode a novel PlexA interacting protein. Figure 5C provides a Northern blot of *Drosophila* total RNA using a probe that included a portion of clone 151.

[0034] Figure 6 provides a series of graphic representations that illustrate the generation and characterization of MICAL Loss-of-Function mutants. Figure 6A Schematic of the screen to remove the *MICAL* locus by generating a small deletion between two P elements that flank *MICAL*. Figure 6B provides a table of summarizing genetic complementation analyses of lines exhibiting the *stretch* wing phenotype. Figure 6C provides a diagram that summarizes complementation analyses and genetic organization of the *MICAL* locus. Sizes are in kilobases (kb); non-continuous sequence is indicated by "//". Figure 6D provides a

Western blot that illustrates that Df(3R)swp2^{MICAL} is a *MICAL* null allele that produces no MICAL protein. Prominent bands are observed at 530kD, 330kD, 300kD, 200kDa, and 125kDa in wild type and at stronger intensity in *MICAL* duplication embryo; none of these bands are observed in *Df(3R)swp2^{MICAL}* embryos. Arrows indicate bands predicted from *MICAL* cDNA analysis (see text; Figure 1A).

[0035] Figure 7 identifies various domains of the *Drosophila* MICAL medium variant polypeptide (SEQ ID NO:10). Flavoprotein Monooxygenase domain is indicated by a squiggly underline; Calponin Homology domain is indicated by gray highlighting; MICAL Homology Region of Unknown Function is indicated by italics; LIM Domain is indicated by black highlighting; Proline Rich (Putative SH3 Ligands) are indicated by single underlining; Putative IQ (calcium) Binding Domain is indicated by dashed underlining; Putative Ena (Ena-like Proteins) binding Domain (Renfranz and Beckerle, *Curr. Opin. Cell. Biol.* 14:88, (2002)) (proline rich region) is indicated by underlining and italics; Plexin Interacting Region is indicated by bold; PDZ Ligand is indicated by double underline.

[0036] Figure 8 provides a diagram that indicates the various domains of MICAL polypeptides and their amino acid residue numbers. Note: The MICAL 2 and 3 plexin interacting region is numbered backwards due to the lack of the intervening sequence denoted "….." .

[0037] Figure 9 provides a diagram that indicates the various domains of MICAL-Like polypeptides and their amino acid residue numbers.

[0038] Figure 10 provides identifies various domains of the mouse MICAL-1 polypeptide (SEQ ID NO:21). Flavoprotein Monooxygenase domain is indicated by a squiggly underline; Calponin Homology domain is indicated by gray highlighting; MICAL Homology Region of Unknown Function is indicated by italics; LIM Domain is indicated by black highlighting; Proline Rich (Putative SH3 Ligands) are indicated by single underlining; Putative IQ (calcium) Binding Domain is indicated by dashed underlining; Putative Ena (Ena-like Proteins) (Renfranz and Beckerle, *Curr. Opin. Cell. Biol.* 14:88, (2002)) binding Domain (proline rich region) is indicated by underlining and italics; Plexin Interacting Region is indicated by bold; PDZ Ligand is indicated by double underline.

[0039] Figure 11 provides the cDNA (SEQ ID NO:1) and encoded polypeptide (SEQ ID NO:2) sequence of human MICAL-1.

[0040] Figure 12 provides the cDNA (SEQ ID NO:3) and encoded polypeptide (SEQ ID NO:4) sequence of human MICAL-2.

[0041] Figure 13 provides the cDNA (SEQ ID NO:5) and encoded polypeptide (SEQ ID NO:6) sequence of human MICAL-3.

[0042] Figure 14 provides the cDNA (SEQ ID NO:7) and encoded polypeptide (SEQ ID NO:8) sequence of the *Drosophila* MICAL long isoform.

[0043] Figure 15 provides the cDNA (SEQ ID NO:9) and encoded polypeptide (SEQ ID NO:10) sequence of the *Drosophila* MICAL medium isoform.

[0044] Figure 16 provides the cDNA (SEQ ID NO:11) and encoded polypeptide (SEQ ID NO:12) sequence of the *Drosophila* MICAL short isoform.

[0045] Figure 17 provides the cDNA (SEQ ID NO:13) and encoded polypeptide (SEQ ID NO:14) sequence of human MICAL-Like 1.

[0046] Figure 18 provides the cDNA (SEQ ID NO:15) and encoded polypeptide (SEQ ID NO:16) sequence of human MICAL-Like 2.

[0047] Figure 19 provides the cDNA (SEQ ID NO:17) and encoded polypeptide (SEQ ID NO:18) sequence of *Drosophila* MICAL-Like.

[0048] Figure 20 provides MICALs in other species. The amino acid sequence through the Flavoprotein Monooxygenase Domain is shown for the species indicated. The numbers indicate the amino acid number for which it aligns with *Drosophila* MICA medium isoform)(e.g., 53 aligns to *Drosophila* MICAL amino acid 53). Percent amino acid identity to the corresponding region of *Drosophila* MICAL is shown.

DETAILED DESCRIPTION OF THE INVENTION

[0049] The present invention is based on the identification of a family of flavoprotein monooxygenases. This family of proteins is involved in the regulation of repulsive axon guidance. While not wanting to be limited by a particular theory, as illustrated herein, this protein family appears to regulate repulsive axon guidance by directly associating with plexins. Through this association, "MICALS" appear to be required for semaphorin-mediated repulsive axon guidance. Furthermore, MICAL proteins contain multiple domains that are known to be important for interactions with actin, intermediate filaments, and cytoskeletal-associated adaptor proteins. Therefore, MICALs are excellent candidates for directly mediating the cytoskeletal alterations characteristic of semaphorin signaling and provide novel targets for the attenuation of axonal repulsion.

[0050] In one embodiment, the present invention provides an isolated polypeptide that includes one or more of an N-terminal MICAL domain, a calponin homology domain, a LIM domain, a proline rich region, and a plexin interacting region, wherein the polypeptide has monooxygenase activity, plexin interacting activity, and/or axon guidance regulatory activity. The polypeptide can also include a first variable MICAL region and a second variable MICAL region, typically surrounding the LIM domain. The second variable region in certain aspects forms a portion of the proline rich region and can include the LIM domain. Accordingly, in one aspect, the polypeptide is a MICAL polypeptide. A MICAL polypeptide includes the following domain organization from N-terminal to C-terminal: an N-terminal MICAL domain, a calponin homology domain, a first variable MICAL region, a LIM domain, a proline rich region, and a plexin interacting region. Furthermore, a MICAL polypeptide has monooxygenase activity and interacts with a plexin, typically plexin A.

[0051] The polypeptide can be a mammalian MICAL polypeptide. For example, the isolated polypeptide can be human MICAL-1, human MICAL-2, or human MICAL-3. Accordingly, the polypeptide can include an amino acid sequence as set forth in SEQ ID NO:2 (human MICAL-1), SEQ ID NO:4 (human MICAL-2), or SEQ ID NO:6 (human MICAL 3).

[0052] The isolated polypeptide can be a *Drosophila* MICAL polypeptide. For example, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:8 (*Drosophila* long variant), SEQ ID NO:10 (*Drosophila* long variant), or SEQ ID NO:12 (*Drosophila* long variant).

[0053] MICALs are also referred to as 151 proteins or Zephyrins. The arrangement of domains within a typical MICAL polypeptide are shown in FIG. 1A. As indicated above, a MICAL polypeptide of the present invention typically includes the following domain organization from N-terminal to C-terminal: an N-terminal MICAL domain, a calponin homology domain, a first variable MICAL region, a LIM domain, a proline rich region, and a plexin interacting region. Furthermore, a MICAL polypeptide has monooxygenase activity and interacts with a plexin, typically plexin A. The MICALs appear unique with respect to containing both calponin homology (CH) and LIM domains, in addition to their conserved N- and C-terminal regions (Fig. 2A).

[0054] In certain aspects, the present invention provides a polypeptide that includes a calponin homology domain. In fact, a MICAL polypeptide of the present invention includes a calponin homology domain. A calponin homology domain is a domain that has at least 30% amino acid sequence identity to the calponin homology domain of SEQ ID NO:2, residues 508 to 612, SEQ ID NO:4, residues 516 to 622, SEQ ID NO:6, residues 518 to 624, SEQ ID NO:8, residues 562 to 669, SEQ ID NO:10, residues 562 to 669, and/or SEQ ID NO:12, residues 562 to 669. In certain aspects, the polypeptide through the calponin homology domain can interact with actin. The calponin homology domain in certain aspects has at least 40%, 50%, 70%, 75%, 80%, 90%, 95%, or 99% sequence identity to SEQ ID NO:2, residues 508 to 612, SEQ ID NO:4, residues 516 to 622, SEQ ID NO:6, residues 518 to 624, SEQ ID NO:8, residues 562 to 669, SEQ ID NO:10, residues 562 to 669, and/or SEQ ID NO:12, residues 562 to 669.

[0055] A polypeptide of the present invention in certain aspects includes a LIM domain. In fact, A MICAL polypeptide of the present invention includes a LIM domain. A LIM domain (Bach (2000), *supra*) is a domain that has at least 30% amino acid sequence identity to a LIM domain of SEQ ID NO:2, residues 697 to 750, SEQ ID NO:4, residues 1002 to 1056, SEQ ID NO:6, residues 792 to 851, SEQ ID NO:8, residues 1074 to 1129, SEQ ID

NO:10, residues 1074 to 1129, and/or SEQ ID NO:12, residues 806 to 861. The LIM domain in certain aspects has at least 40%, 50%, 70%, 75%, 80%, 90%, 95%, or 99% sequence identity to the LIM domain of SEQ ID NO:2, residues 697 to 750, SEQ ID NO:4, residues 1002 to 1056, SEQ ID NO:6, residues 792 to 851, SEQ ID NO:8, residues 1074 to 1129, SEQ ID NO:10, residues 1074 to 1129, and/or SEQ ID NO:12, residues 806 to 861. LIM domains mediate protein/protein interactions with other LIM domain-containing proteins (See Bach (2000), *supra*).

[0056] MICAL polypeptide of the present invention includes a proline rich region. The proline rich region includes the proline rich region indicated in figure 1B as well as variable region 2 indicated in Figure 1B. Accordingly, the proline rich region extends between the LIM domain and the Plexin-interacting region, and includes a variable proline rich domain and a conserved proline rich domain. Thus, the proline rich region extends from the first residue of the N terminal of the Plexin interacting domain to the last C terminal residue of the LIM domain. The proline rich region is defined by the PXXP motifs (the SH3 binding domains).A proline rich region is a region that has at least 1 PXXP SH3 binding domain. In certain aspects, the proline rich region has at least 5, 6, 7, 8, 9, 10, 12, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 SH3 binding domains. The proline rich region in certain aspects of the invention has at least 1%, 2%, 3%, 4%, 5%, 6%, or 7% proline residues. The proline rich region in certain aspects has at least 40%, 50%, 70%, 75%, 80%, 90%, 95%, or 99% sequence identity to the proline rich region of SEQ ID NO:2, residues 751-909, SEQ ID NO:4, residues 1057 to a residue that is 230 amino acids from the C terminus, SEQ ID NO:6, residues 852 to a residue that is 190 amino acids from the C terminus, SEQ ID NO:8, residues 1130-4522, SEQ ID NO:10, residues 1130-2801, and/or SEQ ID NO:12, residues 862-2533. The proline rich region includes numerous potential SH3 binding domains (See e.g., Wages et al., J. Virol. 66(4):1866-74 (1992)) and can include at least one Ena binding domain. For example, there are 18 putative SH3 binding domains in the medium isoform (within the variable region 2) of Drosophila MICAL.

[0057] A MICAL polypeptide of the present invention includes a plexin interacting region at its C-terminus. This domain is typically immediately C-terminal to the proline rich region. Typically, the plexin interacting region contains a predicted heptad-repeat,

coiled-coil structure (Figure 1B), a motif thought to be involved in protein-protein interactions (Burkhard et al., 2001). Interestingly, this region of a MICAL of the present invention typically shares amino acid similarity with several other coiled-coil domain-containing proteins including a portion of the alpha domain found in the Ezrin, Radixin, and Moesin (ERM) proteins (~22% identity; Bretscher et al., 2000).

[0058] A plexin interacting region is a region that has at least 30% amino acid sequence identity to the plexin interacting region of SEQ ID NO:2, residues 910 to 1067, SEQ ID NO:4, residues 348 to 509 after the missing intervening sequence (labeled "..."), SEQ ID NO:6, residues 800 to 989 after the missing intervening sequence (labeled "..."), SEQ ID NO:8, residues 4522 to 4723, SEQ ID NO:10, residues 2802 to 3002, and/or SEQ ID NO:12, residues 2534 to 2734. The proline rich region in certain aspects has at least 40%, 50%, 70%, 75%, 80%, 90%, 95%, or 99% sequence identity to the proline rich region of SEQ ID NO:2, residues 910 to 1067, SEQ ID NO:4, residues 348 to 509 after the missing intervening sequence (labeled "..."), SEQ ID NO:6, residues 800 to 989 after the missing intervening sequence (labeled "..."), SEQ ID NO:8, residues 4522 to 4723, SEQ ID NO:10, residues 2802 to 3002, and/or SEQ ID NO:12, residues 2534 to 2734.

[0059] In certain aspect, the last four amino acids of MICAL (ESII) are a PDZ protein binding motif (Harris and Lim, 2001).

[0060] Typically, MICAL polypeptides of the present invention have two regions of varying length (See e.g., Figure 1B), a first variable MICAL region and a second variable MICAL region, that have no significant similarity to any other proteins, and that appear to determine the size of the different MICAL proteins (Figure 1B). The second variable region includes a high concentration of proline residues, and as indicated above, forms a portion of the proline rich region. For example, the second variable region of Drosophila MICAL medium isoform has 124 proline residues out of 1663 (i.e., 7.5% proline). The variable region and the proline rich region in figure 1B, for the Drosophila medium isoform has 130 prolines out of 1671 residues (i.e., 7.8% proline).

[0061] Interposed between the first and the second variable regions, MICALs typically have a LIM domain as discussed above (Figure 1B), a protein-protein interaction module

found in a variety of proteins involved in signal transduction cascades and in cytoskeletal organization (Bach, 2000), and also a calponin homology (CH) domain as discussed above (Figure 1B), a domain also found in cytoskeletal and signal transduction proteins and known to be involved in actin filament binding (Gimona et al., 2002).

[0062] The present invention also provides an isolated polypeptide as disclosed above, wherein the polypeptide includes an N-terminal MICAL domain having at least about 40%, 45%, 50%, 55%, 60%, 65%. 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to the N-terminal MICAL domains of SEQ ID NOS:2, 4, 6, 8, 10, or 12, and has monooxygenase activity.

[0063] The MICAL N-terminal domain of ~500 amino acids is highly conserved among MICAL-related proteins, but is unique over its entire length in comparison to other proteins. This N-terminal region is referred to herein as the N-terminal MICAL domain. In certain aspects, the N-terminal MICAL domain includes the portion of a MICAL polypeptide that is N-terminal to the Calponin Homology domain. The N-terminal domain can be, for example about 500 to about 561 amino acids in length. N-terminal MICAL domains include residues 1-484 in SEQ ID NO:2, residues 1-492 in SEQ ID NO:4, residues 1-492 in SEQ ID NO:6, and residues 44-529 of SEQ ID Nos:8, 10, and 12.

[0064] The N-terminal MICAL domain typically includes a consensus dinucleotide binding sequence, GxGxxG (Figures 1B and 2A) which is distinct from the sequence present in classical mononucleotide binding motifs (Eggink et al., 1990; Eppink et al., 1997; Schulz, 1992; Wierenga et al., 1986). The N-terminal MICAL domain also typically includes three separate sequence motifs spaced throughout this domain that define them as flavoprotein monooxygenases (also called hydroxylases), a subclass of oxidoreductases (Eggink et al., 1990; Eppink et al., 1997; Wierenga et al., 1986). The amino acid sequence surrounding the GXGXXG motif typically match the consensus sequence for the ADP binding region of flavin adenine dinucleotide (FAD) binding proteins (Rossmann fold or FAD Fingerprint 1, Figures 1B and 2A), and distinguishes this region from consensus NAD, or NADP binding folds (Vallon, 2000; Wierenga et al., 1986). The N-terminal MICAL domain also typically has a well-conserved GD motif (FAD Fingerprint 2; Figures 1B and 2A) C-terminal to the FAD Fingerprint 1 region, which is important for binding the ribose

moiety of FAD (Eggink et al., 1990; Eppink et al., 1997). Finally, N-terminal MICAL domains typically have the conserved DG motif ("Conserved Motif"; Figures 1B and 2A) between the FAD Fingerprint 1 and 2 motifs that has been reported to be involved in binding the pyrophosphate moiety of FAD (Eppink et al., 1997).

[0065] Proteins with the consensus FAD binding regions of MICALs bind FAD and use FAD in the catalysis of oxidation-reduction reactions. Flavoprotein monooxygenases are oxidoreductases (enzymes that catalyze oxidation and reduction reactions) and catalyze the insertion of one atom of molecular oxygen into their substrate using nucleotides as electron donors (Massey, 1995). Accordingly, MICAL polypeptides of the present invention typically have flavoprotein monooxygenase activity. Like other monooxygenases of this type, polypeptides of the present invention typically use FAD as a co-enzyme. Apart from the three consensus regions reported above, monooxygenases typically vary significantly, reflecting the wide range of enzymes in this family and their variable substrate binding pockets also encompassed within this domain (Eppink et al., 1997). However, MICALs and other monooxygenases show significant similarity within these three FAD binding regions and also similar spacing of these regions within the monooxygenase domain. Therefore, polypeptides of the present invention typically have a high degree of identity with the three FAD binding regions (e.g., greater than 50%, 75%, 80%, 85%, 90%, 95%. 98% and 99% sequence identity) and similar spacing of these regions.

[0066] As indicated above, polypeptides of the present invention have monooxygenase activity and bind FAD through their N-terminal MICAL domain. Methods are known in the art for determining whether a polypeptide binds FAD. For example, a solution of a polypeptide of the present invention, including a solution of a polypeptide that includes the N-terminal MICAL domain but not the other domains typically present on MICALs, such as MICAL-FM (see the Examples section herein), is yellow in color, a characteristic of flavoproteins.

[0067] Accordingly, spectral analysis of a polypeptide of the present invention can be used to further identify the polypeptide as flavin binding polypeptides, and therefore to further determine whether a polypeptide includes an N-terminal MICAL domain. For example, a polypeptide that includes an N-terminal MICAL domain, especially a

polypeptide that includes none of the other MICAL domains, typically has distinctive spectral properties. For example, a polypeptide that includes the N-terminal MICAL domain but none of the other MICAL domains can show an absorption peak at around 450 nm and a shoulder at around 360 nm (Figure 3B). This is similar to the absorption spectra of FAD itself (~450nm and ~360nm; Macheroux, 1999), and to other related flavoproteins (e.g., p-Hydroxybenzoate Hydroxylase, Hosokawa and Stanier, 1966; and GidA, White et al., 2001). Monooxygenase activity and other flavoprotein activities can also be determined using other standard protocols as listed in Flavoprotein Protocols, *Methods in Molecular Biology*, Vol. 131, S.K. Chapman and G.A. Reid (Eds), Humana Press, 1999, which is incorporated by reference herein in its entirety.

[0068] A MICAL polypeptide of the present invention can include additional functional domains observed on MICAL polypeptides (See Fig. 7). These additional domains include for example, a MICAL homology region of unknown function (Fig. 7), a putative IQ (calcium) binding domain and a PDZ ligand. Accordingly, in certain aspects, a MICAL polypeptide of the present invention includes, optionally from N-terminal to C-terminal, a flavoprotein monooxygenase domain (N-terminal MICAL domain), a calponin homology domain, a MICAL homology region of unknown function; a LIM Domain; a proline rich region (Putative SH3 Ligands), a putative IQ (calcium) binding domain; a putative Ena binding domain (Ena-like Proteins) (Renfranz and Beckerle, *Curr. Opin. Cell. Biol.* 14:88, (2002)), and a plexin interacting region.

[0069] A MICAL polypeptide of the present invention can include a vimentin interacting region.

[0070] As discussed in more detail herein, the disclosed sequences provided in this specification can be used to identify MICALs or MICAL-Like polypeptides of other species. For example, the sequences disclosed herein were used to search public EST and genomic databases to identify the sequence of mouse MICAL-1 (SEQ ID NO:21) (FIG. 10) (see below). Accordingly, in another aspect, the present invention provides a mouse MICAL polypeptide, such as a polypeptide according to SEQ ID NO:21, SEQ ID NO:22, or SEQ ID NO:23.

[0071] Using the *Drosophila* MICAL amino acid sequence (or MICALs from human or any other species) the publicly available EST and genomic databases can be searched to identify MICAL proteins in other species. The MICAL N terminal flavoprotein monooxygenase domain identified herein, is extremely highly conserved between MICALs in all species (>50% identity). Furthermore, the N terminal flavoprotein monooxygenase domain (i.e., the N-terminal MICAL domain) is not very similar (<<<20% identity) to anything else in the data base (in the genome) over the MICAL ~500 amino acid flavoprotein monooxygenase stretch. Therefore, using this domain to search is extremely fast and accurate to find MICAL sequences. The searching strategy can be a standard, routinely used search strategy. The degree of similarity between MICALs in all species is so high that by searching a public database with the *Drosophila* MICAL flavoprotein monooxygenase amino acid sequence using the publicly available Blast searching program's TBLASTN command one can pull out the coding exons from ordered or unordered (rough draft) genomic sequence. TBLASTN searches a nucleotide database for DNA that is translated by the program into amino acid residues. The alignments can be observed on a web browser. The aligned sequence can then be assembled to identify the coding sequence for the MICAL(s) for any species of interest. An EST nucleotide database can also be searched in a similar manner to a genomic database and as quickly. The sequences provided herein, including the sequences of Figure 20 and the mouse MICAL sequences (SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23), provide examples of sequences that were obtained in this manner. This strategy would effectively and unequivocally identify all of the raw sequences in the database that code for a MICAL protein.

[0072] Furthermore, the rest of the MICAL protein is also highly conserved. Therefore, the other portions of the *Drosophila* MICAL amino acid sequence and the amino acid sequence from those portions of the flavoprotein monooxygenase domain previously identified for a particular species, can then be used to assemble publicly available sequences to identify the full MICAL protein. This can efficiently be done by searching similar databases and using a similar strategy and piecing together the aligned sequence to get a full MICAL sequence.

[0073] In certain aspects, a polypeptide of the present invention is at least 100, 200, 300, 400, 500, 1000, 1500, 2000, 2500, or 3000 amino acids in length.

[0074] In another aspect, a polypeptide of the present invention is a functional portion of a MICAL polypeptide. A functional portion of a MICAL polypeptide is a polypeptide that includes at least an N-terminal MICAL domain that retains monooxygenase activity and/or a functional MICAL plexin interacting region.

[0075] In another embodiment, the present invention provides an isolated polypeptide that includes an N-terminal MICAL domain, but not one or more of the other domains and regions typically found on a MICAL polypeptide. In one aspect, the polypeptide has an N-terminal MICAL domain but no other domain or region of a MICAL polypeptide. The N-terminal MICAL domain for example, is at least 40%, 50%, 75%, 80%, 90%, 95%, 98%, 99%, or 100% identical to an N-terminal MICAL domain of a naturally-occurring MICAL, such as *Drosophila* MICAL or human MICAL 1, 2, or 3. The polypeptide typically retains monooxygenase activity and typically includes the consensus dinucleotide binding sequence and three motifs found in flavoprotein monooxygenases disclosed above.

[0076] The MICAL polypeptides of the present invention have axon guidance regulatory activity. Axon guidance regulatory activity is the ability to affect the positioning, steering, and/or outgrowth of an axon *in vivo* or *in vitro*. Not to be limited by theory, it is believed that MICAL polypeptides of the present invention regulate axon guidance by associating with plexins, thereby being involved in semaphorin-plexin mediated repulsive axon guidance, especially Semaphorin 1a (Sema-1a)-PlexA-mediated repulsive axon guidance, as discussed in more detail hereinbelow.

[0077] The examples section herein illustrates several methods that can be used to identify axon guidance regulatory activity, referred to herein as axon guidance regulatory activity assays. For example, where the polypeptide being analyzed, or an ortholog thereof, is encoded for by a *Drosophila* gene, *Drosophila* mutants can be generated that are loss of function or gain of function mutants. For example, by deleting all or part of the gene encoding the polypeptide, a loss of function mutant can be generated. If the polypeptide has axon guidance regulatory activity, *Drosophila* loss of function mutants devoid of the

function of the polypeptide should exhibit motor axon guidance defects similar to the distinct and highly penetrant defects seen in Sema1a and PlexA loss of function mutants and seen in the MICAL loss of function mutants discussed in the Examples section.

[0078] As another example, axon guidance regulatory activity can be identified by employing an in vitro rat DRG growth cone repulsion assay (Messersmith et al., 1995). The method involves co-culturing E14/15 rat DRG explants with 293 or COS cells expressing Sema3A in the presence of an inhibitor of an on-test polypeptide, as illustrated in the Examples section. NGF-dependent DRG axons exhibit little to no outgrowth toward Sema3A-secreting 293 cell aggregates. If the on-test polypeptide has axon guidance activity then inhibitors of the activity of the polypeptide will inhibit axon repulsion (See e.g., Fig. 4). Axon guidance regulatory activity can also be determined, for example, using single cell turning assays as described by Poo et al. (*Neuron*, 19, 1225-35 (1997)), growth cone collapse assays as described by Raper et al. (See e.g., Luo et al, *Cell* 75, 217-27 (1993)), and mouse knock-out genetic approaches where phenotypes can be observed that must originate from loss of a repulsive response (based on expression data of the ligand, etc…) (See e.g., Giger et al., *Neuron*, 25, 29 (2000)).

[0079] It will be recognized that strategies used herein to identify MICALs and MICAL-like proteins can be used to identify additional MICAL polypeptides, such as MICAL polypeptides of other mammalian species. For example, as illustrated in the Examples, a yeast 2 hybrid system that uses the terminal highly conserved "C2" portion of the PlexA cytoplasm domain can be used to screen cDNA libraries prepared from any organism. Additionally, MICAL polypeptides can be identified by the ability to rescue mutant organisms, such as mutant *Drosophila* prepared using methods disclosed herein, which lack MICAL function. Finally, recombinant DNA technologies can be used to identify and/or develop polynucleotides that encode MICAL or MICAL-Like polypeptides, that are related to, but distinct from those disclosed herein as discussed in more detail hereinbelow.

[0080] Accordingly, the present invention provides an isolated polypeptide as disclosed above, wherein the isolated polypeptide is at least, for example, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 99.9% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

[0081] A polypeptide of the present invention can be a MICAL variant, ortholog, isoform, or mutant. Provided herein are sequences of certain human and *Drosophila* MICALs. However, it will be recognized that additional variants are likely to exist in a population, which have different sequences than the disclosed MICALs. These alleles will likely be highly related in sequence to the disclosed MICALs, and can be further identified as a MICAL by their axonal guidance regulation activity, and by the position in the genome of the polynucleotide sequence encoding the allele. Therefore, based on the presently disclosed MICAL sequences, orthologous sequences in other species besides human or *Drosophila*, such as rat or mouse, can be identified using methods well known in the art, as illustrated herein. Methods for identifying MICAL polynucleotides of other species are discussed in detail below.

[0082] Furthermore, additional MICAL polypeptides and polynucleotides encoding the MICAL polypeptides, can be identified using protein alignment tools, such as those reported in the Examples included herein. As will be recognized, these tools include, for example, PFAM, BLAST, PRINTS, JALVIEW, AND ClustalX, some of which are discussed in more detail herein.

[0083] Finally, as disclosed herein, MICAL transcripts in at least some species are alternately spliced so as to give rise to different polypeptide isoforms from the same MICAL gene. These polypeptide variants or isoforms are examples of MICAL polypeptides of the present invention. MICAL genes cover greater than 40kb of genomic sequences and have at least 25 exons (See Fig. 1A and Fig. 5). Based on analysis of isolated cDNAs and Western analysis (See Fig. 6D), there are at least three *Drosophila* MICAL isoforms, "long" (See e.g. SEQ ID NO:8, "medium" (SEQ ID NO:10), and "short" (SEQ ID NO:12) (Fig. 1A and Figs. 13-15).

[0084] In another embodiment, the present invention provides a MICAL consensus polypeptide. A MICAL consensus polypeptide is a polypeptide that includes at least 50%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% of the most prevalent amino acid occurrences of all MICALs, or of MICALs of a species, such as human MICALs. It will be recognized that amino acid sequences of polypeptides of a protein family such as MICALs, can be aligned and an amino acid sequence of a polypeptide with axonal guidance regulatory

activity can be identified which is different than all of the naturally-occurring family members, but which includes at least some of the most common amino acid occurrences of the naturally-occurring family members.

[0085] Experimental results presented in the Examples section herein demonstrate that at least some MICALs directly associate with plexins and are required for semaphorin-mediated repulsive axon guidance. Furthermore, as disclosed above, MICALs contain multiple domains that are known to be important for interactions with actin, intermediate filaments, and cytoskeletal-associated adaptor proteins. Therefore MICALs are excellent candidates for directly mediating the cytoskeletal alterations characteristic of semaphorin signaling and provide novel targets for the attenuation of axonal repulsion.

[0086] During neural development axons reach their appropriate targets by interpreting a myriad of guidance cues present in their environment. Semaphorin proteins, one of the largest families of guidance cues, are known to influence axon pathfinding, fasciculation, branching, and neuronal cell migration (He et al., 2002; Raper, 2000). A chemorepulsive role in axon guidance has been extensively demonstrated both *in vitro* and *in vivo* for many semaphorins, but they also mediate attractive neuronal guidance.

[0087] The 7 classes of semaphorins include both transmembrane and secreted proteins and are evolutionarily conserved, structurally and in many cases functionally, from invertebrates to vertebrates (Semaphorin Nomenclature Committee, 1999). For example, the transmembrane semaphorin Sema-1a in *Drosophila* is present on developing motor axons and acts as a repellent to regulate motor axon fasciculation *in vivo* (Yu et al., 1998). The related vertebrate transmembrane semaphorin Sema6A also functions as a repellent for axons of sympathetic neurons *in vitro* (Xu et al., 2000). Sema 3A, a well-characterized vertebrate secreted semaphorin, is a potent axonal repellent for a variety of neurons *in vitro*, and *in vivo* serves as a chemorepellent essential for the establishment of many axonal pathways (Raper, 2000). Similarly, the related *Drosophila* secreted semaphorin Sema-2a is expressed on developing muscles and regulates motor axon pathfinding as a target-derived chemorepellent (Matthes et al., 1995; Winberg et al., 1998a).

[0088] Insight into how semaphorins signal repulsive guidance comes from work showing that plexins, a large family of evolutionarily conserved transmembrane proteins, serve as signal transducing receptors for both membrane-bound and secreted semaphorins (Tamagnone and Comoglio, 2000). The four classes of plexins have been found to associate directly with members of five different semaphorin classes. In the *Drosophila* nervous system plexin A (PlexA) is a functional receptor *in vivo* for Sema1a-mediated motor axon repulsion (Winberg et al., 1998b). In vertebrates, repulsion mediated by class 3 secreted semaphorins is dependent on plexin function both *in vitro* and *in vivo* (Cheng et al., 2001; reviewed in Tamagnone and Comoglio, 2000). However, repulsive guidance mediated by class 3 semaphorins, including Sema3A and Sema3F, requires a holoreceptor complex which includes a ligand-binding obligate co-receptor, neuropilin-1 or neuropilin-2, and a class A plexin. Plexin cytoplasmic domains are highly conserved and, for certain A class plexins, are responsible for signaling semaphorin-mediated repulsive axon guidance (Cheng et al., 2001; Takahashi and Strittmatter, 2001).

[0089] The repulsive nature of semaphorin signaling mediated by plexin receptors is due to the modification of the growth cone cytoskeleton. For example, following exposure to secreted Sema3A, growth cones undergo rapid collapse which is accompanied by the depolymerization of F-actin and decreased ability to polymerize new F-actin (Fan et al., 1993). Several modulators of cytoskeletal dynamics have been implicated in this process including Rho family GTPases, p21-activated kinase (PAK), and LIM kinase (Liu and Strittmatter, 2001; Whitford and Ghosh, 2001). In addition, members of the collapsin response mediator protein (CRMP) family, the Ig superfamily protein L1, intracellular levels of cGMP, and the catalytically inactive receptor tyrosine kinase family member offtrack (OTK), have also been implicated in transducing semaphorin repulsive guidance (He et al., 2002). It remains unknown, however, how plexins directly regulate the activity of these signaling molecules in order to modulate cytoskeletal dynamics..

[0090] In another embodiment, the present invention provides an isolated polypeptide that includes a plexin interacting region. The plexin interacting region is typically at least 40%, 50%, 75%, 80%, 90%, 95%, 98%, 99%, or 100% identical to a plexin interacting region of a naturally-occurring MICAL, such as *Drosophila* MICAL or human MICAL 1, 2,

or 3, and retains the ability to interact with a plexin. The plexin interacting region for example, can be at least 90% identical to a plexin interacting region of *Drosophila* MICAL (SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12), or to a plexin interacting region of human MICAL 1 (SEQ ID NO:2), human MICAL 2 (SEQ ID NO:4), human MICAL 3 (SEQ ID NO:6), or a conservative variant thereof. A conservative variant is a polypeptide that is identical to another polypeptide except for conservative amino acid substitutions, as discussed hereinbelow.

[0091] A polypeptide of this embodiment of the invention typically retains the ability to specifically interact with all or part of a plexin either directly or indirectly. For example, the isolated polypeptide can directly interact with the C2 domains of a PlexA, such as PlexA3 and PlexA4. As disclosed in the Examples herein, human MICAL-1 and mouse MICAL-2 specifically interact with the C2 domains of human PlexA3 and mouse PlexA4, respectively. Indirect interactions can be identified, for example, using genetic approaches illustrated in the Examples herein.

[0092] Methods for determining whether an on-test polypeptide is capable of interacting with a plexin are well-known in the art. For example, traditional methods of identifying specific protein interactions can be used. Accordingly, immunoprecipitation can be used to identify whether a polypeptide interacts with a plexin by determining whether the on-test polypeptide and a plexin coimmunoprecipitate, as illustrated in the Examples section. Furthermore, for example, a plexin protein can be isolated on a protein gel, and binding of a labeled on-test polypeptide can be determined. Alternatively, for example, a yeast interaction assay can be used, as disclosed in the Examples herein (See Fig. 5).

[0093] The isolated polypeptide that includes a plexin interacting region, in certain aspects, does not have one or more other domains and/or activities typically present in a MICAL polypeptide. For example, the polypeptide in certain aspects does not have monooxygenase activity. The polypeptide of this embodiment of the invention can be a mutant MICAL polypeptide that acts as a dominant negative mutant with respect to MICAL activity. For example, as illustrated in the Examples section, the polypeptide can be a truncated MICAL polypeptide that includes at least one, but not all, functional domains

typically present on a MICAL polypeptide. Alternatively, the mutant can include mutations that alter, for example by destroying, certain MICAL functions.

[0094] The examples section provided herein provides a MICAL$^{G \rightarrow W}$ mutant (SEQ ID NO:20) that is mutated in the three glycine residues within the FAD fingerprint 1 motif of MICAL to tryptophan, a mutation known in other proteins to disrupt FAD binding without altering the overall structure of the protein (Kubo et al., 1997; Lawton and Philpot, 1993; Wierenga et al., 1986). As illustrated in the Examples section, the MICAL$^{G \rightarrow W}$ mutant which includes an intact plexin-interacting domain but is functionally inactive, exerts a dominant-negative effect on motor axon guidance in a wild-type genetic background. Not to be limited by theory, it is believed that MICAL$^{G \rightarrow W}$ exerts its dominant negative effect by binding to a plexin, thereby competing for binding of wild-type MICAL to the plexin target.

[0095] As illustrated in the Examples herein, a polypeptide according to this embodiment of the invention can be a truncated mutant that only includes a plexin-interacting region. Accordingly, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:19. The polypeptide can be targeted to the membrane by including a membrane targeting sequence such as an N-terminal myristoylation sequence as illustrated in the Examples section herein (see mutant MICAL$^{Myr \rightarrow CT}$). Other membrane targeting sequences such as, for example, a palmitylation sequence, can be used.

[0096] A polypeptide of this embodiment of the invention, can include, for example, a MICAL or a MICAL-Like plexin interacting domain, since both of these protein families include a plexin interacting domain.

[0097] In another embodiment the present invention relates to a family of MICAL-like (MICAL-L) proteins, members of which have a similar organization to MICALs but lack the region N-terminal to the CH domain (Figure 2B). MICAL-L proteins include at least one MICAL-L protein in *Drosophila* (D-MICAL-L) and at least two family members in humans. D-MICAL-L cDNA and genomic DNA sequence information suggest that D-MICAL-L Plexin interacting domain begins just N-terminal to the CH domain of a MICAL protein. Analysis of publicly available mammalian cDNA and genomic sequences suggests that human MICAL-L1 and MICAL-L2 are similar in overall domain organization to D-

MICAL-L and do not contain the highly conserved ~500 amino acid MICAL N-terminal domain.

[0098] Accordingly, the present invention provides an isolated polypeptide that includes a plexin interacting region and alternatively one or more of a calponin homology domain, a LIM domain, and a proline rich region. The polypeptide can also include a first variable region and a second variable region. Accordingly, in one aspect, the polypeptide is or includes a MICAL-like polypeptide. A MICAL-Like polypeptide includes, from N-terminal to C-terminal, a calponin homology domain, a first variable MICAL region, a LIM domain, a second variable MICAL region, a proline rich region, and a plexin interacting region.

[0099] A polypeptide according to this aspect of the invention typically specifically interacts with a plexin, as discussed above for MICAL polypeptides of the present invention.

[0100] The polypeptide, for example, includes a calponin homology domain, followed by a first variable region, followed by a LIM domain, followed by a second variable region, followed by a proline rich region, and followed by a plexin interacting region. Such polypeptides include a variant, ortholog, isoform, or mutant of a MICAL-Like protein disclosed herein.

[0101] In one aspect, the polypeptide is a mammalian MICAL-Like polypeptide. For example, the isolated polypeptide can be human MICAL-Like 1 or human MICAL-Like 2. Accordingly, the isolated polypeptide can have an amino acid sequence that is at least 40%, 50%, 75%, 80%, 90%, 95%, 98%, 99%, or 100% identical to a an amino acid sequence as set forth in SEQ ID NO:14 (human MICAL-Like 1) or SEQ ID NO:16 (human MICAL-like 2).

[0102] In another aspect, the isolated polypeptide is a *Drosophila* MICAL-L polypeptide. For example, the polypeptide can have an amino acid sequence as set forth in SEQ ID NO:18 (*Drosophila* MICAL-Like), or a variant, ortholog, isoform, or mutant thereof.

[0103] In another aspect, the polypeptide of the present invention is a functional portion of a MICAL-Like polypeptide. A functional portion of a MICAL polypeptide is a polypeptide that includes at least a functional domain, for example a functional MICAL plexin interacting region.

[0104] A functional peptide portion of a MICAL or MICAL-Like polypeptide for example, can be obtained by examining peptide portions of a MICAL or MICAL-Like polypeptide using methods as provided herein or other standard methods, to identify fragments that retain at least one of the activities of a wild-type MICAL including the ability to interact with a plexin, particularly Plexin A, and monooxygenase activity.

[0105] A functional peptide portion of a MICAL or MICAL-Like polypeptide that specifically interacts with plexin can be identified using any of various assays known to be useful for identifying specific protein-protein interactions. Such assays include, for example, methods of gel electrophoresis, affinity chromatography, the two hybrid system of Fields and Song (*Nature* 340:245-246, 1989; see, also, U.S. Patent No. 5,283,173; Fearon et al., *Proc. Natl. Acad. Sci., USA* 89:7958-7962, 1992; Chien et al., *Proc. Natl. Acad. Sci. USA* 88:9578-9582, 1991; Young, *Biol. Reprod.* 58:302-311(1998), each of which is incorporated herein by reference), the reverse two hybrid assay (Leanna and Hannink, *Nucl. Acids Res.* 24:3341-3347, 1996, which is incorporated herein by reference), the repressed transactivator system (U.S. Patent No. 5,885,779, which is incorporated herein by reference), the phage display system (Lowman, *Ann. Rev. Biophys. Biomol. Struct.* 26:401-424, 1997, which is incorporated herein by reference), GST/HIS pull down assays, mutant operators (WO 98/01879, which is incorporated herein by reference), the protein recruitment system (U.S. Patent No. 5,776,689, which is incorporated herein by reference), and the like (see, for example, Mathis, *Clin. Chem.* 41:139-147, 1995 Lam, *Anticancer Drug Res.* 12:145-167, 1997; Phizicky et al., *Microbiol. Rev.* 59:94-123, 1995; each of which is incorporated herein by reference).

[0106] A functional peptide portion of a MICAL or MICAL-Like polypeptide also can be identified using methods of molecular modeling. For example, an amino acid sequence of a MICAL or MICAL-Like polypeptide can be entered into a computer system having appropriate modeling software, and a three dimensional representation of the MICAL or

MICAL-Like polypeptide ("virtual MICAL" or "virtual MICAL-Like polypeptide ") can be produced. A MICAL or MICAL-Like polypeptide amino acid sequence also can be entered into the computer system, such that the modeling software can simulate portions of the MICAL or MICAL-Like polypeptide sequence, and can identify those peptide portions that can interact specifically, for example, with the virtual plexin.

[0107] It should be recognized that such methods, including two hybrid assays and molecular modeling methods, also can be used to identify other specifically interacting molecules encompassed within the present invention. For example, the methods can be used to identify other proteins to which MICALS and/or MICAL-Like proteins bind, as revealed by the various domains of these proteins.

[0108] Modeling systems useful for the purposes disclosed herein can be based on structural information obtained, for example, by crystallographic analysis or nuclear magnetic resonance analysis, or on primary sequence information (see, for example, Dunbrack et al., "Meeting review: the Second meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2) (Asilomar, California, December 13-16, 1996). *Fold Des.* 2(2): R27-42, (1997); Fischer and Eisenberg, *Protein Sci.* 5:947-55, 1996; (see, also, U.S. Patent No. 5,436,850); Havel, *Prog. Biophys. Mol. Biol.* 56:43-78, 1991; Lichtarge et al., *J. Mol. Biol.* 274:325-37, 1997; Matsumoto et al., *J. Biol. Chem.* 270:19524-31, 1995; Sali et al., *J. Biol. Chem.* 268:9023-34, 1993; Sali, *Molec. Med. Today* 1:270-7, 1995a; Sali, *Curr. Opin. Biotechnol.* 6:437-51, 1995b; Sali et al., *Proteins* 23: 318-26, 1995c; Sali, *Nature Struct. Biol.* 5:1029-1032, 1998; U.S. Patent No. 5,933,819; U.S. Patent No. 5,265,030, each of which is incorporated herein by reference).

[0109] The crystal structure coordinates of a MICAL or MICAL-Like polypeptide can be used to design compounds that bind to the protein and alter its physical or physiological properties in a variety of ways. The structure coordinates of the protein can also be used to computationally screen small molecule databases for agents that bind to the polypeptide to develop modulating or binding agents, which can act as agonists or antagonists of MICAL axon guidance regulatory activity. Such agents can be identified by computer fitting kinetic data using standard equations (see, for example, Segel, "Enzyme Kinetics" (J. Wiley & Sons 1975), which is incorporated herein by reference).

[0110] Methods of using crystal structure data to design inhibitors or binding agents are known in the art. For example, MICAL or MICAL-Like polypeptide coordinates can be superimposed onto other available coordinates of similar proteins, including proteins having a bound inhibitor, to provide an approximation of the way the inhibitor interacts with the receptor. Computer programs employed in the practice of rational drug design also can be used to identify compounds that reproduce interaction characteristics similar to those found, for example, between a MICAL or MICAL-Like polypeptide and a co-crystallized plexin. Detailed knowledge of the nature of the specific interactions allows for the modification of compounds to alter or improve solubility, pharmacokinetics, and the like, without affecting binding activity.

[0111] Computer programs for carrying out the activities necessary to design agents using crystal structure information are well known. Examples of such programs include, Catalyst Databases™ - an information retrieval program accessing chemical databases such as BioByte Master File, Derwent WDI and ACD; Catalyst/HYPO™ - generates models of compounds and hypotheses to explain variations of activity with the structure of drug candidates; Ludi™ - fits molecules into the active site of a protein by identifying and matching complementary polar and hydrophobic groups; and Leapfrog™ - "grows" new ligands using a genetic algorithm with parameters under the control of the user.

[0112] Various general purpose machines can be used with such programs, or it may be more convenient to construct more specialized apparatus to perform the operations. Generally, the embodiment is implemented in one or more computer programs executing on programmable systems each comprising at least one processor, at least one data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. The program is executed on the processor to perform the functions described herein.

[0113] Each such program can be implemented in any desired computer language, including, for example, machine, assembly, high level procedural, or object oriented programming languages, to communicate with a computer system. In any case, the language may be a compiled or interpreted language. The computer program will typically be stored on

a storage media or device, for example, a ROM, CD-ROM, magnetic or optical media, or the like, that is readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

[0114] Embodiments of the invention include systems, for example, internet based systems, particularly computer systems which store and manipulate coordinate information obtained by crystallographic or NMR analysis, or amino acid or nucleotide sequence information, as disclosed herein. As used herein, the term "computer system" refers to the hardware components, software components, and data storage components used to analyze coordinates or sequences as set forth herein. The computer system typically includes a processor for processing, accessing and manipulating the sequence data. The processor can be any well known type of central processing unit, for example, a Pentium II or Pentium III processor from Intel Corporation, or a similar processor from Sun, Motorola, Compaq, Advanced MicroDevices or International Business Machines.

[0115] Typically the computer system is a general purpose system that comprises the processor and one or more internal data storage components for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

[0116] Where it is desired to identify a chemical entity that interacts specifically with MICAL or MICAL-Like polypeptide, any of several methods to screen chemical entities or fragments for their ability to interact specifically with the molecule can be used. This process may begin by visual inspection, for example, of MICAL or MICAL-Like polypeptide on the computer screen. Selected peptide portions of MICAL or MICAL-Like polypeptides, or chemical entities that can act as mimics, then can be positioned in a variety of orientations, or docked, within an individual binding site of the MICAL or MICAL-Like polypeptides. Docking can be accomplished using software such as Quanta and Sybyl, followed by energy

minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER.

[0117] Specialized computer programs can be particularly useful for selecting peptide portions of a prodomain, or chemical entities useful, for example, as a MICAL or MICAL-Like polypeptide agonist or antagonist. Such programs include, for example, GRID (Goodford, *J. Med. Chem.*, 28:849-857, 1985; available from Oxford University, Oxford, UK); MCSS (Miranker and Karplus, *Proteins: Structure. Function and Genetics* 11:29-34, 1991, available from Molecular Simulations, Burlington MA); AUTODOCK (Goodsell and Olsen, *Proteins: Structure. Function, and Genetics* 8:195-202, 1990, available from Scripps Research Institute, La Jolla CA); DOCK (Kuntz, et al., *J. Mol. Biol.* 161:269-288, 1982, available from University of California, San Francisco CA), each of which is incorporated herein by reference.

[0118] Suitable peptides or agents that have been selected can be assembled into a single compound or binding agent. Assembly can be performed by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen, followed by manual model building using software such as Quanta or Sybyl. Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include, for example, CAVEAT (Bartlett et al, *Special Pub., Royal Chem. Soc.* 78:182-196, 1989, available from the University of California, Berkeley CA); 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro CA; for review, see Martin, *J. Med. Chem.* 35:2145-2154, 1992); HOOK (available from Molecular Simulations, Burlington, Mass.), each of which is incorporated herein by reference.

[0119] In another embodiment, the present invention provides an isolated polynucleotide encoding a polypeptide of the present invention disclosed hereinabove. Accordingly, a polynucleotide of the present invention can encode a MICAL polypeptide or a MICAL-like polypeptide of the present invention. A polynucleotide of the present invention that encodes a MICAL polypeptide encodes an isolated polypeptide that includes one or more of an N-terminal MICAL domain, a calponin homology domain, a LIM domain, a proline rich region, and a plexin interacting region, wherein the polypeptide has monooxygenase activity, plexin interacting activity, and/or axon guidance regulatory activity. The encoded

polypeptide can also include a first variable region and a second variable region surrounding the LIM domain.

[0120] A polynucleotide of the present invention that encodes a MICAL-Like polypeptide encodes an isolated polypeptide that includes one or more of a calponin homology domain, a LIM domain, a proline rich region, and a plexin interacting region, wherein the polypeptide has plexin interacting activity. The encoded polypeptide can also include a first variable region and a second variable region surrounding the LIM domain.

[0121] In one aspect the polynucleotide encodes a mammalian MICAL polypeptide, or a functional portion thereof, or MICAL-like polypeptide, or a functional portion thereof. For example, the polynucleotide can encode all or a portion of human MICAL-1, human MICAL-2, human MICAL-3, human MICAL-Like 1, or human MICAL-like 2. As such the polynucleotide can include all or a portion (e.g. a cDNA, or a coding region) of a human MICAL-1 gene, human MICAL-2 gene, human MICAL-3 gene, human MICAL-Like 1 gene, or human MICAL-like 2 gene. The polynucleotide, for example, can include a coding region or an entire transcript. Accordingly, the polynucleotide can encode a polypeptide that includes an amino acid sequence as set forth in SEQ ID NO:2 (human MICAL-1), SEQ ID NO:4 (human MICAL-2), SEQ ID NO:6 (human MICAL-3), SEQ ID NO:14 (human MICAL-Like 1), or SEQ ID NO:16 (human MICAL-Like 2), or an isoform thereof.

[0122] The polynucleotide can include a polynucleotide that is at least 50%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to a MICAL coding sequence as set forth in figures 11 to 16. Accordingly, a polynucleotide of the present invention in certain aspects, includes a coding nucleotide portion of SEQ ID NO:1 (human MICAL-1 cDNA), SEQ ID NO:3 (human MICAL-2 coding sequence), SEQ ID NO:5 (human MICAL-3 cDNA), SEQ ID NO:13 (human MICAL-Like 1 cDNA), or SEQ ID NO:14 (human MICAL-Like 2 cDNA), or a portion thereof. The polynucleotide can include an entire MICAL cDNA or gene, or an entire MICAL-Like cDNA or gene, or a portion thereof.

[0123] A polynucleotide according to this embodiment of the invention can encode a *Drosophila* MICAL polypeptide or MICAL-Like polypeptide, for example a polypeptide having the sequence as set forth in SEQ ID NO:8 (*Drosophila* MICAL long isoform), SEQ

ID NO:10 (*Drosophila* MICAL medium isoform), SEQ ID NO:12 (*Drosophila* MICAL short isoform), or SEQ ID NO:18 (*Drosophila* MICAL-Like polypeptide). For example, the polynucleotide can be 50%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to a nucleotide sequence as set forth in SEQ ID NO:7 (Drosophila MICAL long isoform cDNA sequence), SEQ ID NO:9 (Drosophila MICAL medium isoform cDNA sequence), SEQ ID NO:11 (Drosophila MICAL short isoform cDNA sequence), or SEQ ID NO:17 (Drosophila MICAL-Like cDNA sequence), or a portion thereof.

[0124] Polynucleotides of the present invention are typically at least 15, 25, 50, 75, 100, 125, 150, 200, 250, 500, 1000, 2500, 5000, 10000, 25000, 5000, 10000, 15000, 20000, 25000, 30000, or 40,000 nucleotides in length.

[0125] In another embodiment, the present invention provides a polynucleotide that specifically hybridizes to a polynucleotide that encodes a MICAL polypeptide or a MICAL-Like polypeptide of the present invention. For example, the polynucleotide can specifically hybridize to a polynucleotide that encodes all or a portion of a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18, or a complement thereof, that in certain aspects has monooxygenase activity.

[0126] A polynucleotide of the present invention can inhibit expression of a polynucleotide that encodes a MICAL polypeptide or a MICAL-Like polypeptide of the present invention. The polynucleotide can include a polynucleotide that is complementary to a nucleotide sequence that is at least 50%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to a all or a portion, such as a coding portion, of a nucleotide sequence as set forth in SEQ ID NO:1 (human MICAL-1 cDNA), SEQ ID NO:3 (human MICAL-2 cDNA), SEQ ID NO:5 (human MICAL-3 cDNA), SEQ ID NO:13 (human MICAL-like 1 cDNA), or SEQ ID NO:15 (human MICAL-like 2 cDNA), or a portion thereof.

[0127] Polynucleotides encoding MICAL or MICAL-Like polypeptides of various organisms in addition to those identified herein, can be identified using well known procedures and algorithms based on identity (or homology) to the disclosed sequences. Homology or identity is often measured using sequence analysis software such as the

Sequence Analysis Software Package of the Genetics Computer Group (University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various deletions, substitutions and other modifications. The terms "homology" and "identity," when used herein in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or of nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region as measured using any number of sequence comparison algorithms or by manual alignment and visual inspection.

[0128] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0129] The term "comparison window" is used broadly herein to include reference to a segment of any one of the number of contiguous positions, for example, about 20 to 600 positions, for example, amino acid or nucleotide position, usually about 50 to about 200 positions, more usually about 100 to about 150 positions, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm of Smith and Waterman (Adv. Appl. Math. 2:482, 1981), by the homology alignment algorithm of Needleman and Wunsch (J. Mol. Biol. 48:443, 1970), by the search for similarity method of Person and Lipman (Proc. Natl. Acad. Sci., *USA* 85:2444, 1988), each of which is incorporated herein by reference; by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI); or by manual alignment and visual inspection. Other algorithms for

determining homology or identity include, for example, in addition to a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLocks IMProved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome databases to identify polynucleotide sequences having substantially identical sequences.

[0130] A number of genome databases are available for comparison through the National Center for Biotechnology internet site (http://www.ncbi.nlm.nih.gov/). For example, the NCBI site provides access to the complete genomes of human (Ventor, J.C., et al., *Science* 291:1304-1351 (2001), *M. genitalium*, *M. jannaschii*, *H. influenzae*, *E. coli*, yeast (*S. cerevisiae*), and *D. melanogaster (*Adams et al., *Science* 287: 2185-2195 (2000)).

[0131] One example of a useful algorithm is BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (*Nucleic Acids Res.* 25:3389-3402, 1977; *J. Mol. Biol. 215*:403-410, 1990, each of which is incorporated herein by reference). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*, 1977, 1990). These initial

neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, *Proc. Natl. Acad. Sci., USA* 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0132] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, for example, Karlin and Altschul, Proc. Natl. Acad. Sci., *USA* 90:5873, 1993, which is incorporated herein by reference). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0133] In one embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST"). In particular, five specific BLAST programs are used to perform the following task:

[0134] (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

[0135] (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

[0136] (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

[0137] (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

[0138] (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

[0139] The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., Science 256:1443-1445, 1992; Henikoff and Henikoff, Proteins 17:49-61, 1993, each of which is incorporated herein by reference). Less preferably, the PAM or PAM250 matrices may also be used (Schwartz and Dayhoff, eds., "Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure" (Washington, National Biomedical Research Foundation 1978)). BLAST programs are accessible through the U.S. National Library of Medicine, for example, at www.ncbi.nlm.nih.gov.

[0140] The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some embodiments, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

[0141] Therefore, using MICAL sequences disclosed herein, and known methods and databases, a polynucleotide encoding a MICAL or MICAL-Like polypeptide, or the MICAL or MICAL-Like polypeptide or protein can be identified from any organism. Therefore, MICAL polynucleotides, polypeptides, and proteins of the present invention include, for

example, mouse, rat, cow, pig, horse, dog, human, chicken, turkey, zebrafish, and other species.

[0142] It should also be recognized that reference is made herein to particular peptides or polypeptides beginning or ending at "about" a particular amino acid residue. The term "about" is used in this context because it is recognized that a particular protease can cleave a MICAL polypeptide at or immediately adjacent to a proteolytic cleavage recognition site, or one or a few amino acids from the recognition site. As such, reference, for example, to a MICAL polypeptide having a sequence of about amino acid residues 1 to 263 of SEQ ID NO:2 would include an amino terminal peptide portion of MICAL that has a carboxy terminus ending at amino acid residue 257 to amino acid residue 269, preferably at amino acid residue 260 to amino acid residue 266.

[0143] The term "peptide," "peptide portion," or polypeptide is used broadly herein to mean two or more amino acids linked by a peptide bond. The term "fragment" or "proteolytic fragment" also is used herein to refer to a product that can be produced by a proteolytic reaction on a polypeptide, i.e., a peptide produced upon cleavage of a peptide bond in the polypeptide. Although the term "proteolytic fragment" is used generally herein to refer to a peptide that can be produced by a proteolytic reaction, it should be recognized that the fragment need not necessarily be produced by a proteolytic reaction, but also can be produced using methods of chemical synthesis or methods of recombinant DNA technology, as discussed in greater detail below, to produce a synthetic peptide that is equivalent to a proteolytic fragment. In view of the disclosed homology of MICALs and MICAL-Like proteins with other proteins, it will be recognized that a polypeptide of the invention is characterized, in part, in that it is not present in previously disclosed members of this superfamily. Whether a polypeptide portion of a MICAL or MICAL-Like polypeptide is present in a previously disclosed protein readily can be determined using the computer algorithms described above.

[0144] Generally, a peptide or polypeptide of the invention contains at least about six amino acids, usually contains about ten amino acids, and can contain fifteen or more amino acids, particularly twenty or more amino acids. It should be recognized that the terms "peptide" and “polypeptide” is not used herein to suggest a particular size or number of

amino acids comprising the molecule, and that a polypeptide of the invention can contain up to several amino acid residues or more.

[0145] As used herein, the term "substantially purified" or "substantially pure" or "isolated" means that the molecule being referred to, for example, a polypeptide or a polynucleotide, is in a form that is relatively free of proteins, nucleic acids, lipids, carbohydrates or other materials with which it is naturally associated. Generally, a substantially pure polypeptide, polynucleotide, or other molecule constitutes at least twenty percent of a sample, generally constitutes at least about fifty percent of a sample, usually constitutes at least about eighty percent of a sample, and particularly constitutes about ninety percent or ninety-five percent or more of a sample. A determination that a peptide or a polynucleotide of the invention is substantially pure can be made using well known methods, for example, by performing electrophoresis and identifying the particular molecule as a relatively discrete band. A substantially pure polynucleotide, for example, can be obtained by cloning the polynucleotide, or by chemical or enzymatic synthesis. A substantially pure peptide or polypeptide can be obtained, for example, by a method of chemical synthesis, or using methods of protein purification, followed by proteolysis and, if desired, further purification by chromatographic or electrophoretic methods.

[0146] A polypeptide of the invention can be identified by comparison to a MICAL or MICAL-Like sequence and determining that the amino acid sequence of the polypeptide is contained within the MICAL or MICAL-Like polypeptide sequence, respectively. It should be recognized, however, that a polypeptide of the invention need not be identical to a corresponding amino acid sequence of MICAL or a MICAL-Like polypeptide. Thus, a polypeptide of the invention can correspond to an amino acid sequence of a MICAL polypeptide, for example, but can vary from a naturally occurring sequence, for example, by containing one or more D-amino acids in place of a corresponding L-amino acid; or by containing one or more amino acid analogs, for example, an amino acid that has been derivatized or otherwise modified at its reactive side chain. Similarly, one or more peptide bonds in the polypeptide can be modified. In addition, a reactive group at the amino terminus or the carboxy terminus or both can be modified. Such polypeptides can be modified, for example, to have improved stability to a protease, an oxidizing agent or other

reactive material the peptide may encounter in a biological environment, and, therefore, can be particularly useful in performing a method of the invention. Of course, the polypeptides can be modified to have decreased stability in a biological environment such that the period of time the polypeptide is active in the environment is reduced.

[0147] The sequence of a MICAL or MICAL-Like polypeptide of the invention also can be modified by incorporating a conservative amino acid substitution for one or a few amino acids in the polypeptide. Conservative amino acid substitutions include the replacement of one amino acid residue with another amino acid residue having relatively the same chemical characteristics, for example, the substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, for example, substitution of arginine for lysine; or of glutamic for aspartic acid; or of glutamine for asparagine; or the like. Examples of positions of a MICAL polypeptide that can be modified are evident from examination of differences in the disclosed MICAL sequences.

[0148] The present invention also provides a substantially purified proteolytic fragment of a MICAL polypeptide or a functional peptide portion thereof. A peptide portion of a MICAL polypeptide that is equivalent to a proteolytic fragment of a MICAL can be produced by a chemical method or a recombinant DNA method.

[0149] The term "polynucleotide" is used broadly herein to mean a sequence of two or more deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. As such, the term "polynucleotide" includes RNA and DNA, which can be a gene or a portion thereof, a cDNA, a synthetic polydeoxyribonucleic acid sequence, or the like, and can be single stranded or double stranded, as well as a DNA/RNA hybrid. Furthermore, the term "polynucleotide" as used herein includes naturally occurring nucleic acid molecules, which can be isolated from a cell, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR). In various embodiments, a polynucleotide of the invention can contain nucleoside or nucleotide analogs, or a backbone bond other than a phosphodiester bond (see above).

[0150] In general, the nucleotides that make up a polynucleotide are naturally occurring deoxyribonucleotides, such as adenine, cytosine, guanine or thymine linked to 2'-deoxyribose, or ribonucleotides such as adenine, cytosine, guanine or uracil linked to ribose. However, a polynucleotide also can contain nucleotide analogs, including non-naturally occurring synthetic nucleotides or modified naturally occurring nucleotides. Such nucleotide analogs are well known in the art and commercially available, as are polynucleotides containing such nucleotide analogs (Lin et al., *Nucl. Acids Res.* 22:5220-5234 (1994); Jellinek et al., *Biochemistry* 34:11363-11372 (1995); Pagratis et al., *Nature Biotechnol.* 15:68-73 (1997), each of which is incorporated herein by reference).

[0151] The covalent bond linking the nucleotides of a polynucleotide generally is a phosphodiester bond. However, the covalent bond also can be any of numerous other bonds, including a thiodiester bond, a phosphorothioate bond, a peptide-like bond or any other bond known to those in the art as useful for linking nucleotides to produce synthetic polynucleotides (see, for example, Tam et al., *Nucl. Acids Res.* 22:977-986 (1994); Ecker and Crooke, *BioTechnology* 13:351360 (1995), each of which is incorporated herein by reference). The incorporation of non-naturally occurring nucleotide analogs or bonds linking the nucleotides or analogs can be particularly useful where the polynucleotide is to be exposed to an environment that can contain a nucleolytic activity, including, for example, a tissue culture medium or upon administration to a living subject, since the modified polynucleotides can be less susceptible to degradation.

[0152] A polynucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally will be chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template (Jellinek et al., *supra*, 1995).

[0153] Where a polynucleotide encodes a polypeptide, for example, a polypeptide portion of a MICAL or a polypeptide agent, the coding sequence generally is contained in a

vector and is operatively linked to appropriate regulatory elements, including, if desired, a tissue specific promoter or enhancer. The encoded peptide can be further operatively linked, for example, to a peptide tag such as a His-6 tag or the like, which can facilitate identification of expression of the agent in the target cell. A polyhistidine tag peptide such as His-6 can be detected using a divalent cation such as nickel ion, cobalt ion, or the like. Additional peptide tags include, for example, a FLAG epitope, which can be detected using an anti-FLAG antibody (see, for example, Hopp et al., *BioTechnology* 6:1204 (1988); U.S. Patent No. 5,011,912, each of which is incorporated herein by reference); a c-myc epitope, which can be detected using an antibody specific for the epitope; biotin, which can be detected using streptavidin or avidin; and glutathione S-transferase, which can be detected using glutathione. Such tags can provide the additional advantage that they can facilitate isolation of the operatively linked peptide or peptide agent, for example, where it is desired to obtain a substantially purified peptide corresponding to a proteolytic fragment of a MICAL or MICAL-Like polypeptide.

[0154] As used herein, the term "operatively linked" or "operatively associated" means that two or more molecules are positioned with respect to each other such that they act as a single unit and effect a function attributable to one or both molecules or a combination thereof. For example, a polynucleotide sequence encoding a polypeptide of the invention can be operatively linked to a regulatory element, in which case the regulatory element confers its regulatory effect on the polynucleotide similarly to the way in which the regulatory element would effect a polynucleotide sequence with which it normally is associated with in a cell. A first polynucleotide coding sequence also can be operatively linked to a second (or more) coding sequence such that a chimeric polypeptide can be expressed from the operatively linked coding sequences. The chimeric polypeptide can be a fusion polypeptide, in which the two (or more) encoded peptides are translated into a single polypeptide, i.e., are covalently bound through a peptide bond; or can be translated as two discrete peptides that, upon translation, can operatively associate with each other to form a stable complex.

[0155] A polynucleotide of the invention, including a polynucleotide agent useful in performing a method of the invention, can be contacted directly with a target cell. For

example, oligonucleotides useful as antisense molecules, ribozymes, or triplexing agents can be directly contacted with a target cell, whereupon they enter the cell and affect their function. A polynucleotide agent also can interact specifically with a polypeptide, for example, a MICAL polypeptide, thereby altering the ability of the MICAL to interact specifically with a plexin. Such polynucleotides, as well as methods of making and identifying such polynucleotides, are disclosed herein or otherwise well known in the art (see, for example, O'Connell et al., *Proc. Natl. Acad. Sci., USA* 93:5883-5887, 1996; Tuerk and Gold, *Science* 249:505-510, 1990; Gold et al., *Ann. Rev. Biochem.* 64:763-797, 1995; each of which is incorporated herein by reference).

[0156] A polynucleotide of the invention, which can encode a MICAL or MICAL-Like polypeptide or can encode a mutant MICAL or MICAL-Like polypeptide or functional peptide portion thereof, or can be a polynucleotide agent useful in performing a method of the invention, can be contained in a vector, which can facilitate manipulation of the polynucleotide, including introduction of the polynucleotide into a target cell. The vector can be a cloning vector, which is useful for maintaining the polynucleotide, or can be an expression vector, which contains, in addition to the polynucleotide, regulatory elements useful for expressing the polynucleotide and, where the polynucleotide encodes a peptide, for expressing the encoded peptide in a particular cell. An expression vector can contain the expression elements necessary to achieve, for example, sustained transcription of the encoding polynucleotide, or the regulatory elements can be operatively linked to the polynucleotide prior to its being cloned into the vector.

[0157] An expression vector, or the polynucleotide included on the vector generally contains or encodes a promoter sequence, which can provide constitutive or, if desired, inducible or tissue specific or developmental stage specific expression of the encoding polynucleotide, a poly-A recognition sequence, and a ribosome recognition site or internal ribosome entry site, or other regulatory elements such as an enhancer, which can be tissue specific. The vector also can contain elements required for replication in a prokaryotic or eukaryotic host system or both, as desired. Such vectors, which include plasmid vectors and viral vectors such as bacteriophage, baculovirus, retrovirus, lentivirus, adenovirus, vaccinia virus, semliki forest virus and adeno-associated virus vectors, are well known and can be

purchased from a commercial source (Promega, Madison WI; Stratagene, La Jolla CA; GIBCO/BRL, Gaithersburg MD) or can be constructed by one skilled in the art (see, for example, *Meth. Enzymol.*, Vol. 185, Goeddel, ed. (Academic Press, Inc., 1990); Jolly, *Canc. Gene Ther.* 1:51-64, 1994; Flotte, *J. Bioenerg. Biomemb.* 25:37-42, 1993; Kirshenbaum et al., *J. Clin. Invest.* 92:381-387, 1993; each of which is incorporated herein by reference).

[0158] A tetracycline (tet) inducible promoter can be particularly useful for driving expression of a polynucleotide of the invention, for example, a polynucleotide encoding a dominant negative form of a MICAL polypeptide. Upon administration of tetracycline, or a tetracycline analog, to a subject containing a polynucleotide operatively linked to a tet inducible promoter, expression of the encoded polypeptide is induced, whereby the polypeptide can effect its activity, for example, whereby a polypeptide agent can reduce or inhibit semaphorin mediated axonal repulsion. Such a method can be used, for example, to induce axon formation after spinal cord injury.

[0159] The polynucleotide also can be operatively linked to tissue specific regulatory element, for example, a neuron specific regulatory element, such that expression of an encoded peptide is restricted to the neurons in an individual, or to neurons in a mixed population of cells in culture. Neuron specific regulatory elements are well known in the art as illustrated in the Examples section (See also, e.g., Nelson, S.B., et al., *Mol Endocrinol.* 14:1509-22 (2000); and Navarro et al., *Gene Ther.* 6:1884-92 (1999)). For example, after a spinal cord injury, a vector that encodes a dominant negative MICAL polypeptide operatively linked to a neuron-specific promoter, can be delivered to the site of spinal cord injury. Expression of the dominant negative mutant in neurons is expected to inhibit MICAL regulatory activity, thereby inhibiting semaphorin-mediated axon repulsion. This inhibition permits axons to regrow and migrate to reach new targets

[0160] Viral expression vectors can be particularly useful for introducing a polynucleotide into a cell, particularly a cell in a subject. Viral vectors provide the advantage that they can infect host cells with relatively high efficiency and can infect specific cell types. For example, a polynucleotide encoding a MICAL polypeptide, or functional peptide portion thereof can be cloned into a baculovirus vector, which then can be used to infect an insect host cell, thereby providing a means to produce large amounts of

the encoded MICAL or MICAL-Like protein. The viral vector also can be derived from a virus that infects cells of an organism of interest, for example, vertebrate host cells such as mammalian, avian or piscine host cells. Viral vectors can be particularly useful for introducing a polynucleotide useful in performing a method of the invention into a target cell. Viral vectors have been developed for use in particular host systems, particularly mammalian systems and include, for example, retroviral vectors, other lentivirus vectors such as those based on the human immunodeficiency virus (HIV), adenovirus vectors, adeno-associated virus vectors, herpesvirus vectors, vaccinia virus vectors, and the like (see Miller and Rosman, *BioTechniques* 7:980-990, 1992; Anderson et al., *Nature* 392:25-30 Suppl., 1998; Verma and Somia, *Nature* 389:239-242, 1997; Wilson, *New Engl. J. Med.* 334:1185-1187 (1996), each of which is incorporated herein by reference).

[0161] When retroviruses, for example, are used for gene transfer, replication competent retroviruses theoretically can develop due to recombination of retroviral vector and viral gene sequences in the packaging cell line utilized to produce the retroviral vector. Packaging cell lines in which the production of replication competent virus by recombination has been reduced or eliminated can be used to minimize the likelihood that a replication competent retrovirus will be produced. All retroviral vector supernatants used to infect cells are screened for replication competent virus by standard assays such as PCR and reverse transcriptase assays. Retroviral vectors allow for integration of a heterologous gene into a host cell genome, which allows for the gene to be passed to daughter cells following cell division.

[0162] A polynucleotide, which can be contained in a vector, can be introduced into a cell by any of a variety of methods known in the art (Sambrook et al., *Molecular Cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press 1989); Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, MD (1987, and supplements through 1995), each of which is incorporated herein by reference). Such methods include, for example, transfection, lipofection, microinjection, electroporation and, with viral vectors, infection; and can include the use of liposomes, microemulsions or the like, which can facilitate introduction of the polynucleotide into the cell and can protect the polynucleotide from degradation prior to its introduction into the cell. The selection of a

particular method will depend, for example, on the cell into which the polynucleotide is to be introduced, as well as whether the cell is isolated in culture, or is in a tissue or organ in culture or *in situ*.

[0163] Introduction of a polynucleotide into a cell by infection with a viral vector is particularly advantageous in that it can efficiently introduce the nucleic acid molecule into a cell *ex vivo* or *in vivo* (see, for example, U.S. Patent No. 5,399,346, which is incorporated herein by reference). Moreover, viruses are very specialized and can be selected as vectors based on an ability to infect and propagate in one or a few specific cell types. Thus, their natural specificity can be used to target the nucleic acid molecule contained in the vector to specific cell types. As such, a vector based on an HIV can be used to infect T cells, a vector based on an adenovirus can be used, for example, to infect respiratory epithelial cells, a vector based on a herpesvirus can be used to infect neuronal cells, and the like. Other vectors, such as adeno-associated viruses can have greater host cell range and, therefore, can be used to infect various cell types, although viral or non-viral vectors also can be modified with specific receptors or ligands to alter target specificity through receptor mediated events.

[0164] Thus, a polynucleotide of the invention can be a naturally occurring, synthetic, or intentionally manipulated polynucleotide. For example, portions of the mRNA sequence can be altered due to alternate RNA splicing patterns or the use of alternate promoters for RNA transcription. As another example, the polynucleotide can be subjected to site directed mutagenesis. The polynucleotide also can be antisense nucleotide sequence. MICAL and MICAL-Like polynucleotides (i.e., polynucleotides that encode a MICAL polypeptide or a MICAL-Like polypeptide) of the invention include sequences that are degenerate as a result of the genetic code. There are 20 natural amino acids, most of which are specified by more than one codon. Therefore, all degenerate nucleotide sequences are included within the invention, provided the amino acid sequence of a MICAL polypeptide or a MICAL-Like encoded by the polynucleotide is functionally unchanged.

[0165] Oligonucleotide portions of a polynucleotide encoding a MICAL polypeptide or a MICAL-Like polypeptide of the invention also are encompassed within the present invention. Such oligonucleotides generally are at least about 15 bases in length, which is

sufficient to permit the oligonucleotide to selectively hybridize to a polynucleotide encoding the MICAL or MICAL-Like polypeptide, and can be at least about 18 nucleotides or 21 nucleotides or more in length. As used herein, the term "selective hybridization" or "selectively hybridize" refers to hybridization under moderately stringent or highly stringent physiological conditions, which can distinguish related nucleotide sequences from unrelated nucleotide sequences.

[0166] In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (for example, relative GC:AT content), and nucleic acid type, i.e., whether the oligonucleotide or the target nucleic acid sequence is DNA or RNA, can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter. Methods for selecting appropriate stringency conditions can be determined empirically or estimated using various formulas, and are well known in the art (see, for example, Sambrook et al., *supra*, 1989).

[0167] An example of progressively higher stringency conditions is as follows: 2X SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2X SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2X SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1X SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, for example, high stringency conditions, or each of the conditions can be used, for example, for 10 to 15 minutes each, in the order listed above, repeating any or all of the steps listed.

[0168] A MICAL or MICAL-Like polypeptide-encoding polynucleotide of the invention can be obtained by any of several methods. For example, the polynucleotide can be isolated using hybridization or computer based techniques, as are well known in the art. These methods include, but are not limited to, 1) hybridization of genomic or cDNA libraries with probes to detect homologous nucleotide sequences; 2) antibody screening of expression libraries to detect cloned DNA fragments with shared structural features; 3) polymerase chain reaction (PCR) on genomic DNA or cDNA using primers capable of annealing to the DNA sequence of interest; 4) computer searches of sequence databases for similar

sequences (see above); 5) differential screening of a subtracted DNA library; and 6) two hybrid assays using, for example, a MICAL polypeptide in one of the hybrids.

[0169] A polynucleotide of the invention, for example, a polynucleotide encoding a MICAL, can be derived from a vertebrate species, including a mammalian, avian, or piscine species, or from an invertebrate species. Screening procedures that rely on nucleic acid hybridization allow the isolation any gene sequence from any organism, provided the appropriate probe is available. Oligonucleotide probes that correspond to a part of the sequence encoding the protein in question can be synthesized chemically. Hybridization is particularly useful in the detection of cDNA clones derived from sources where an extremely low amount of mRNA sequences relating to the polypeptide of interest are present. Thus, by using stringent hybridization conditions directed to avoid nonspecific binding, autoradiographic visualization can be used to identify a specific cDNA clone by the hybridization of the target DNA to an oligonucleotide probe in the mixture that is the complete complement of the target nucleic acid (Wallace et al., *Nucl. Acid Res.*, 9:879, 1981, which is incorporated herein by reference). Alternatively, a subtractive library can be used, thereby eliminating nonspecific cDNA clones.

[0170] When the entire amino acid sequence of a desired polypeptide is not known, the direct synthesis of DNA sequences is not possible and the method of choice is the synthesis of cDNA sequences. Among the standard procedures for isolating cDNA sequences of interest is the formation of cDNA libraries prepared in plasmids or phage, wherein the libraries are derived from reverse transcription of mRNA that is abundant in donor cells having a high level of genetic expression. When used in combination with polymerase chain reaction technology, even rare expression products can be cloned. Where significant portions of the amino acid sequence of the polypeptide are known, the production of labeled single stranded or double stranded DNA or RNA probe sequences duplicating a sequence putatively present in the target cDNA can be employed in hybridization procedures carried out on cloned copies of the cDNA, which have been denatured into a single stranded form (Jay et al., *Nucl. Acid Res.*, 11:2325, 1983, which is incorporated herein by reference).

[0171] A cDNA expression library, such as a lambda gt11 library, can be screened for MICAL or MICAL-Like polypeptides using an antibody specific for a MICAL or MICAL-

Like polypeptide. The antibody can be polyclonal or monoclonal, and can be used to detect expression product indicative of the presence of a MICAL or MICAL-Like polypeptide encoding cDNA. Such an expression library also can be screened with a MICAL or MICAL-Like polypeptide to identify a clone encoding at least a portion of a MICAL or MICAL-Like polypeptide binding domain of a plexin or other protein that interacts with MICAL or the MICAL-Like polypeptide.

[0172] Polynucleotides encoding mutant MICAL and MICAL-Like polypeptides are also encompassed within the invention. An alteration in a polynucleotide encoding a MICAL or MICAL-Like protein can be an intragenic mutation such as point mutation, nonsense (STOP) mutation, missense mutation, splice site mutation or frameshift, or can be a heterozygous or homozygous deletion, and can be a naturally occurring mutation or can be engineered using recombinant DNA methods, for example. Such alterations can be detected using standard methods known to those of skill in the art, including, but not limited to, nucleotide sequence analysis, Southern blot analysis, a PCR based analysis such as multiplex PCR or sequence tagged sites (STS) analysis, or *in situ* hybridization analysis. MICAL and MICAL-Like polypeptides can be analyzed by standard SDS-PAGE, immunoprecipitation analysis, western blot analysis, or the like.

[0173] A polynucleotide encoding a MICAL or a MICAL-Like polypeptide can be expressed *in vitro* by introducing the polynucleotide into a suitable host cell. "Host cells" can be any cells in which the particular vector can be propagated, and, where appropriate, in which a polynucleotide contained in the vector can be expressed. The term "host cells" includes any progeny of an original host cell. It is understood that all progeny of the host cell may not be identical to the parental cell due, for example, to mutations that occur during replication. However, such progeny are included when the term "host cell" is used. Methods of obtaining a host cell that transiently or stably contains an introduced polynucleotide of the invention are well known in the art. In one aspect, the present invention provides host cell that includes a polynucleotide encoding a MICAL polypeptide according to the present invention, operably linked to a heterologous promoter.

[0174] In certain aspects of embodiments of the present invention, a cell is a mammalian cell, for example a human cell.

[0175] A polynucleotide encoding a MICAL or a MICAL-Like polypeptide of the invention can be inserted into a vector, which can be a cloning vector or a recombinant expression vector. The term "recombinant expression vector" refers to a plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of a polynucleotide, particularly, with respect to the present invention, a polynucleotide encoding all or a peptide portion of a MICAL or a MICAL-Like polypeptide. Such expression vectors contain a promoter sequence, which facilitates the efficient transcription of the inserted genetic sequence of the host. The expression vector generally contains an origin of replication, a promoter, as well as specific genes which allow phenotypic selection of the transformed cells. Vectors suitable for use in the present invention include, but are not limited to, the T7-based expression vector for expression in bacteria (Rosenberg, et al., *Gene* 56:125, 1987), the pMSXND expression vector for expression in mammalian cells (Lee and Nathans, *J. Biol. Chem.* 263:3521, 1988) and baculovirus-derived vectors for expression in insect cells. The DNA segment can be present in the vector operably linked to regulatory elements, for example, a promoter, which can be a T7 promoter, metallothionein I promoter, polyhedrin promoter, or other promoter as desired, particularly tissue specific promoters or inducible promoters.

[0176] A polynucleotide sequence encoding a MICAL or a MICAL-Like polypeptide can be expressed in either prokaryotes or eukaryotes. Hosts can include microbial, yeast, insect and mammalian organisms. Methods of expressing polynucleotides having eukaryotic or viral sequences in prokaryotes are well known in the art, as are biologically functional viral and plasmid DNA vectors capable of expression and replication in a host. Methods for constructing an expression vector containing a polynucleotide of the invention are well known, as are factors to be considered in selecting transcriptional or translational control signals, including, for example, whether the polynucleotide is to be expressed preferentially in a particular cell type or under particular conditions (see, for example, Sambrook et al., *supra*, 1989).

[0177] A variety of host cell/expression vector systems can be utilized to express a MICAL or a MICAL-Like polypeptide coding sequence, including, but not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage DNA,

plasmid DNA or cosmid DNA expression vectors; yeast cells transformed with recombinant yeast expression vectors; plant cell systems infected with recombinant virus expression vectors such as a cauliflower mosaic virus or tobacco mosaic virus, or transformed with recombinant plasmid expression vector such as a Ti plasmid; insect cells infected with recombinant virus expression vectors such as a baculovirus; animal cell systems infected with recombinant virus expression vectors such as a retrovirus, adenovirus or vaccinia virus vector; and transformed animal cell systems genetically engineered for stable expression. Where the expressed MICAL or a MICAL-Like polypeptide is post-translationally modified, for example, by glycosylation, it can be particularly advantageous to select a host cell/expression vector system that can effect the desired modification, for example, a mammalian host cell/expression vector system.

[0178] Depending on the host cell/vector system utilized, any of a number of suitable transcription and translation elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, and the like can be used in the expression vector (Bitter et al., *Meth. Enzymol.* 153:516-544, 1987). For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage λ, plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like can be used. When cloning in mammalian cell systems, promoters derived from the genome of mammalian cells, for example, a human or mouse metallothionein promoter, or from mammalian viruses, for example, a retrovirus long terminal repeat, an adenovirus late promoter or a vaccinia virus 7.5K promoter, can be used. Promoters produced by recombinant DNA or synthetic techniques can also be used to provide for transcription of the inserted MICAL or a MICAL-Like polypeptide coding sequence.

[0179] In yeast cells, a number of vectors containing constitutive or inducible promoters can be used (see Ausubel et al., *supra*, 1987, see chapter 13; Grant et al., *Meth. Enzymol.* 153:516-544, 1987; Glover, *DNA Cloning* Vol. II (IRL Press, 1986), see chapter 3; Bitter, *Meth. Enzymol.* 152:673-684, 1987; see, also, *The Molecular Biology of the Yeast Saccharomyces* (Eds., Strathern et al., Cold Spring Harbor Laboratory Press, 1982), Vols. I and II). A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL can be used (Rothstein, *DNA Cloning* Vol. II (*supra*, 1986), chapter 3).

Alternatively, vectors can be used which promote integration of foreign DNA sequences into the yeast chromosome.

[0180] Eukaryotic systems, particularly mammalian expression systems, allow for proper post-translational modifications of expressed mammalian proteins. Eukaryotic cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, phosphorylation, and advantageously, plasma membrane insertion of the gene product can be used as host cells for the expression of a MICAL or MICAL-Like polypeptide, or functional peptide portion thereof.

[0181] Mammalian cell systems which utilize recombinant viruses or viral elements to direct expression can be engineered. For example, when using adenovirus expression vectors, the MICAL or MICAL-Like polypeptide coding sequence can be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. Alternatively, the vaccinia virus 7.5K promoter can be used (Mackett et al., *Proc. Natl. Acad. Sci., USA* 79:7415-7419, 1982; Mackett et al., *J. Virol.* 49:857-864, 1984; Panicali et al., *Proc. Natl. Acad. Sci., USA* 79:4927-4931, 1982). Particularly useful are bovine papilloma virus vectors, which can replicate as extrachromosomal elements (Sarver et al., *Mol. Cell. Biol.* 1:486, 1981). Shortly after entry of this DNA into mouse cells, the plasmid replicates to about 100 to 200 copies per cell. Transcription of the inserted cDNA does not require integration of the plasmid into the host cell chromosome, thereby yielding a high level of expression. These vectors can be used for stable expression by including a selectable marker in the plasmid, such as, for example, the neo gene. Alternatively, the retroviral genome can be modified for use as a vector capable of introducing and directing the expression of the MICAL or MICAL-Like gene in host cells (Cone and Mulligan, *Proc. Natl. Acad. Sci., USA* 81:6349-6353, 1984). High level expression can also be achieved using inducible promoters, including, but not limited to, the metallothionein IIA promoter and heat shock promoters.

[0182] For long term, high yield production of recombinant proteins, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with the MICAL or a MICAL-Like polypeptide encoding cDNA controlled by appropriate expression control elements such as promoter, enhancer,

sequences, transcription terminators, and polyadenylation sites, and a selectable marker. The selectable marker in the recombinant plasmid can confer resistance to the selection, and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci, which, in turn can be cloned and expanded into cell lines. For example, following the introduction of foreign DNA, engineered cells can be allowed to grow for 1 to 2 days in an enriched media, and then are switched to a selective media. A number of selection systems can be used, including, but not limited to, the herpes simplex virus thymidine kinase (Wigler et al., *Cell* 11:223, 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, *Proc. Natl. Acad. Sci., USA* 48:2026, 1982), and adenine phosphoribosyltransferase (Lowy, et al., *Cell* 22:817, 1980) genes can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, et al., *Proc. Natl. Acad. Sci. USA* 77:3567, 1980; O'Hare et al., *Proc. Natl. Acad. Sci., USA* 78: 1527, 1981); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci., USA* 78:2072, 1981); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., *J. Mol. Biol.* 150:1, 1981); and hygro, which confers resistance to hygromycin (Santerre et al., *Gene* 30:147, 1984) genes. Additional selectable genes, including trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan, *Proc. Natl. Acad. Sci., USA* 85:8047, 1988); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, *Curr. Comm. Mol. Biol.* (Cold Spring Harbor Laboratory Press, 1987), also have been described.

[0183] When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate coprecipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors can be used. Eukaryotic cells can also be cotransformed with DNA sequences encoding the MICAL or a MICAL-Like polypeptides of the invention, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma

virus, to transiently infect or transform eukaryotic cells and express the protein. (Gluzman, *Eukaryotic Viral Vectors* (Cold Spring Harbor Laboratory Press, 1982)).

[0184] The invention also provides stable recombinant cell lines, the cells of which express MICAL or a MICAL-Like polypeptides and contain DNA that encodes MICAL or a MICAL-Like polypeptides. Suitable cell types include, but are not limited to, NIH 3T3 cells (murine), C2C12 cells, L6 cells, and P19 cells. C2C12 and L6 myoblasts differentiate spontaneously in culture and form myotubes depending on the particular growth conditions (Yaffe and Saxel, *Nature* 270:725-727, 1977; Yaffe, *Proc. Natl. Acad. Sci., USA* 61:477-483, 1968). P19 is an embryonal carcinoma cell line. Such cells are described, for example, in the Cell Line Catalog of the American Type Culture Collection (ATCC). These cells can be stably transformed using well known methods (see, for example, Ausubel et al., *supra*, 1995, see sections 9.5.1-9.5.6).

[0185] A MICAL or a MICAL-Like polypeptide can be expressed from a recombinant polynucleotide of the invention using inducible or constitutive regulatory elements, as described herein. The desired protein encoding sequence and an operably linked promoter can be introduced into a recipient cell either as a non-replicating DNA (or RNA) molecule, which can either be a linear molecule or a covalently closed circular molecule. Expression of the desired molecule can occur due to transient expression of the introduced sequence, or the polynucleotide can be stably maintained in the cell, for example, by integration into a host cell chromosome, thus allowing a more permanent expression. Accordingly, the cells can be stably or transiently transformed (transfected) cells.

[0186] An example of a vector that can be employed is one which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker can complement an auxotrophy in the host such as leu2, or ura3, which are common yeast auxotrophic markers; can confer a biocide resistance, for example, to an antibiotic or to heavy metal ions such as copper, or the like. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or can be introduced into the same cell by cotransfection.

[0187] The introduced sequence can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a variety of vectors can be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include the ease with which recipient cells that contain the vector can be recognized and selected from those cells that do not contain the vector; the number of copies of the vector desired in a particular host cell; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

[0188] For a mammalian host, several vector systems are available for expression. One class of vectors utilizes DNA elements that provide autonomously replicating extra-chromosomal plasmids derived from animal viruses, for example, a bovine papilloma virus, polyoma virus, adenovirus, or SV40 virus. A second class of vectors includes vaccinia virus expression vectors. A third class of vectors relies upon the integration of the desired gene sequences into the host chromosome. Cells that have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers genes (as described above), which allow selection of host cells that contain the expression vector. The selectable marker gene can be directly linked to the DNA sequences to be expressed, or introduced into the same cell by cotransfection. Additional elements can be included to provide for optimal synthesis of an encoded mRNA or peptide, including, for example, splice signals, transcription promoters or enhancers, and transcription or translation termination signals. cDNA expression vectors incorporating appropriate regulatory elements are well known in the art (see, for example, Okayama, *Mol. Cell. Biol.* 3:280, 1983).

[0189] Once the vector or DNA sequence containing the construct has been prepared for expression, the DNA construct can be introduced into an appropriate host. Various methods can be used for introducing the polynucleotide into a cell, including, for example, methods of transfection or transformation such as protoplast fusion, calcium phosphate precipitation, and electroporation or other conventional techniques, for example, infection where the vector is a viral vector.

[0190] The invention also provides transgenic non-human animals that have cells that constitutively express a recombinant MICAL or a MICAL-Like polypeptide or that have

recombinantly inactivated MICAL or MICAL-Like function. In certain aspects, transgenic animals that constitutively express a recombinant MICAL or MICAL-Like protein can be expressed with a tag sequence that can be used to facilitate immunoprecipitation of the MICAL or MICAL-Like polypeptide. Such transgenic non-human animals can be used for example, to facilitate the identification of agents which bind the MICAL or MICAL-Like polypeptides. Alternatively, the transgenic non-human organism can express a mutant transgenic MICAL in the central nervous system or peripheral nervous system, to identify mutant MICAL polypeptides that affect axonal guidance.

[0191] Accordingly, in one aspect, the transgenic animal is a transgenic non-human organism whose genome includes a transgenic DNA sequence that includes a polynucleotide that encodes a mutant MICAL polypeptide operably linked to a promoter that is active in the central nervous system and/or peripheral nervous system, wherein the mouse expresses the transgenic polynucleotide in the central nervous system and/or peripheral nervous system, and wherein expression levels of transgenic polynucleotide are sufficient to effect an axonal guidance phenotype of the non-human organism.

[0192] In another aspect, the transgenic animal is a non-human transgenic animal having a genome comprising a transgene comprising a nucleotide sequence encoding a MICAL polypeptide operably linked to a heterologous promoter. The non-human transgenic animal expresses the transgenic polynucleotide in the central nervous system and/or peripheral nervous system at expression levels sufficient to effect an axonal guidance phenotype of the non-human organism. In one aspect, the MICAL polypeptide is ectopically expressed. The MICAL polypeptide is expressed in the transgenic animal at a greater level in one or more cells of the non-human transgenic animal than the MICAL polypeptide is expressed in comparable cells of a comparable non-human transgenic animal.

[0193] In another embodiment, the present invention provides a non-human transgenic animal having a genome comprising a recombinantly inactivated nucleotide sequence encoding a MICAL polypeptide that has been recombinantly inactivated. The non-human transgenic animal has an altered phenotype that results from inactivation of the MICAL polypeptide. For example, the altered phenotype can be an altered axon guidance phenotype.

[0194] The non-human transgenic animal of this aspect of the invention, in certain aspects is heterozygous for the nucleotide sequence that has been inactivated. Alternatively, the non-human transgenic animal of this aspect of the invention, can be homozygous for the nucleotide sequence that has been recombinantly inactivated.

[0195] As used herein, the term "transgenic," when used in reference to an animal or an organism, means that cells of the animal or organism have been genetically manipulated to contain an exogenous polynucleotide sequence that is stably maintained with the cells. The manipulation can be, for example, microinjection of a polynucleotide or infection with a recombinant virus containing the polynucleotide. Thus, the term "transgenic" is used herein to refer to animals (organisms) in which one or more cells receive a recombinant polynucleotide, which can be integrated into a chromosome in the cell, or can be maintained as an extrachromosomally replicating polynucleotide, such as might be engineered into a yeast artificial chromosome. The term "transgenic animal" also includes a "germ cell line" transgenic animal. A germ cell line transgenic animal is a transgenic animal in which the genetic information has been taken up and incorporated into a germ line cell, therefore conferring the ability to transfer the information to offspring. If such offspring in fact possess some or all of that information, the offspring also are considered to be transgenic animals. The invention further encompasses transgenic organisms.

[0196] A transgenic organism can be any organism whose genome has been altered by *in vitro* manipulation of an early stage embryo or a fertilized egg, or by any transgenic technology to induce a specific gene knock-out. The term "gene knock-out" refers to the targeted disruption of a gene in a cell or *in vivo* that results in complete loss of function. Gene knock-outs is also referred to herein as inactivated genes, such as recombinantly inactivated genes. A target gene in a transgenic animal can be rendered nonfunctional by an insertion targeted to the gene to be rendered nonfunctional, for example, by homologous recombination, or by any other method for disrupting the function of a gene in a cell.

[0197] The transgene to be used in the practice of the subject invention can be a DNA sequence comprising a modified MICAL or MICAL-Like polypeptide coding sequence. Preferably, the modified MICAL or MICAL-Like gene is one that is disrupted by homologous targeting in embryonic stem cells. For example, the entire MICAL gene can be

deleted (See Examples herein). Optionally, the disruption (or deletion) can be accompanied by insertion of or replacement with another polynucleotide, for example, a polynucleotide encoding a nonfunctional MICAL or MICAL-Like polypeptide. A "knock-out" phenotype also can be conferred by introducing or expressing an antisense MICAL or MICAL-Like polypeptide encoding polynucleotide in a cell in the organism, or by expressing an antibody or a dominant negative MICAL or MICAL-Like polypeptide in the cells.

[0198] Various methods are known for producing a transgenic animal. In one method, an embryo at the pronuclear stage (a "one cell embryo") is harvested from a female and the transgene is microinjected into the embryo, in which case the transgene will be chromosomally integrated into the germ cells and somatic cells of the resulting mature animal. In another method, embryonic stem cells are isolated and the transgene is incorporated into the stem cells by electroporation, plasmid transfection or microinjection; the stem cells are then reintroduced into the embryo, where they colonize and contribute to the germ line. Methods for microinjection of polynucleotides into mammalian species are described, for example, in U.S. Patent No. 4,873,191, which is incorporated herein by reference. In yet another method, embryonic cells are infected with a retrovirus containing the transgene, whereby the germ cells of the embryo have the transgene chromosomally integrated therein.

[0199] When the animals to be made transgenic are avian, microinjection into the pronucleus of the fertilized egg is problematic because avian fertilized ova generally go through cell division for the first twenty hours in the oviduct and, therefore, the pronucleus is inaccessible. Thus, the retrovirus infection method is preferred for making transgenic avian species (see U.S. Patent No. 5,162,215, which is incorporated herein by reference). If microinjection is to be used with avian species, however, the embryo can be obtained from a sacrificed hen approximately 2.5 hours after the laying of the previous laid egg, the transgene is microinjected into the cytoplasm of the germinal disc and the embryo is cultured in a host shell until maturity (Love et al., *Biotechnology* 12, 1994). When the animals to be made transgenic are bovine or porcine, microinjection can be hampered by the opacity of the ova, thereby making the nuclei difficult to identify by traditional differential

interference-contrast microscopy. To overcome this problem, the ova first can be centrifuged to segregate the pronuclei for better visualization.

[0200] Non-human transgenic animals of the invention can be an invertebrate or a vertebrate. For example, the transgenic organism can be *Drosophila* or a mammal such as a mouse or a rat. The transgene can be introduced into embryonal target cells at various developmental stages, and different methods are selected depending on the stage of development of the embryonal target cell. The zygote is the best target for microinjection. The use of zygotes as a target for gene transfer has a major advantage in that the injected DNA can incorporate into the host gene before the first cleavage (Brinster et al., *Proc. Natl. Acad. Sci., USA* 82:4438-4442, 1985). As a consequence, all cells of the transgenic non-human animal carry the incorporated transgene, thus contributing to efficient transmission of the transgene to offspring of the founder, since 50% of the germ cells will harbor the transgene.

[0201] A transgenic animal can be produced by crossbreeding two chimeric animals, each of which includes exogenous genetic material within cells used in reproduction. Twenty-five percent of the resulting offspring will be transgenic animals that are homozygous for the exogenous genetic material, 50% of the resulting animals will be heterozygous, and the remaining 25% will lack the exogenous genetic material and have a wild type phenotype.

[0202] In the microinjection method, the transgene is digested and purified free from any vector DNA, for example, by gel electrophoresis. The transgene can include an operatively associated promoter, which interacts with cellular proteins involved in transcription, and provides for constitutive expression, tissue specific expression, developmental stage specific expression, or the like. Such promoters include those from cytomegalovirus (CMV), Moloney leukemia virus (MLV), and herpes virus, as well as those from the genes encoding metallothionein, skeletal actin, Phosphenolpyruvate carboxylase (PEPCK), phosphoglycerate (PGK), dihydrofolate reductase (DHFR), and thymidine kinase (TK). Promoters from viral long terminal repeats (LTRs) such as Rous sarcoma virus LTR also can be employed. When the animals to be made transgenic are avian, preferred promoters include those for the chicken β-globin gene, chicken lysozyme gene, and avian leukosis

virus. Constructs useful in plasmid transfection of embryonic stem cells will employ additional regulatory elements, including, for example, enhancer elements to stimulate transcription, splice acceptors, termination and polyadenylation signals, ribosome binding sites to permit translation, and the like.

[0203] In the retroviral infection method, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, *Proc. Natl. Acad. Sci, USA* 73:1260-1264, 1976). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan et al., *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al., *Proc. Natl. Acad. Sci., USA* 82:6927-6931, 1985; Van der Putten et al., *Proc. Natl. Acad. Sci, USA* 82:6148-6152, 1985). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus producing cells (Van der Putten et al., *supra*, 1985; Stewart et al., *EMBO J.* 6:383-388, 1987). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al., *Nature* 298:623-628, 1982). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic nonhuman animal. Further, the founder can contain various retroviral insertions of the transgene at different positions in the genome, which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit with low efficiency, by intrauterine retroviral infection of the mid-gestation embryo (Jahner et al., *supra*, 1982).

[0204] Embryonal stem cell (ES) also can be targeted for introduction of the transgene. ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. *Nature* 292:154-156, 1981; Bradley et al., *Nature* 309:255-258, 1984; Gossler et al., *Proc. Natl. Acad. Sci., USA* 83:9065-9069, 1986; Robertson et al., *Nature* 322:445-448, 1986). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter

colonize the embryo and contribute to the germ line of the resulting chimeric animal (see Jaenisch, *Science* 240:1468-1474, 1988).

[0205] The present invention also provides an antibody or antigen binding fragment thereof that specifically bind a MICAL polypeptide or a MICAL-Like polypeptide, or a functional peptide portion thereof. Particularly useful antibodies of the invention include antibodies that specifically bind a plexin interacting region of a MICAL, thereby inhibiting binding of the MICAL to plexins. Such antibodies can be useful, for example, for inhibiting semaphorin-mediate axonal repulsion and thus stimulating, for example, regeneration of axon connections after spinal cord injury. In certain aspects, the present invention provides an antibody or antigen binding fragment that binds an N-terminal MICAL domain, a MICAL calponin homology domain, a MICAL LIM domain, a MICAL proline rich region, or a MICAL plexin-interacting region.

[0206] A monoclonal antibody that binds specifically to a MICAL or MICAL-Like polypeptide can be used to treat a pathological condition involving, for example failure of axon regrowth. In a preferred embodiment, the MICAL or MICAL-Like polypeptide antibody is administered to a patient by intravenous, intramuscular subcutaneous injection, or direct injection to a site of spinal cord damage. A monoclonal antibody can be administered, for example, within a dose range between about 0.1 µg/kg to about 100 mg/kg; more preferably between about 1 µg/kg to 75 mg/kg; most preferably from about 10 mg/kg to 50 mg/kg. The antibody can be administered, for example, by bolus injunction or by slow infusion. Slow infusion over a period of 30 minutes to 2 hours is preferred. The anti- MICAL or anti-MICAL-Like polypeptide antibody, can be formulated in a formulation suitable for administration to a patient. Such formulations are known in the art.

[0207] The dosage regimen will be determined by the attending physician considering various factors which modify the action of the MICAL or MICAL-Like polypeptide protein or the plexin protein, for example, amount of tissue desired to be formed, the site of tissue damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue, the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage can vary with the type of matrix used in the reconstitution and the types of agent, such as anti-MICAL or anti-MICAL-Like polypeptide antibodies, to

be used in the composition. Generally, systemic or injectable administration, such as intravenous, intramuscular, subcutaneous injection, or injection to a site of damage of the nervous system is employed. Administration generally is initiated at a dose which is minimally effective, and the dose is increased over a preselected time course until a positive effect is observed. Subsequently, incremental increases in dosage are made limiting such incremental increases to such levels that produce a corresponding increase in effect, while taking into account any adverse affects that can appear. The addition of other agents that promote neuron process regrowth, can also affect the dosage.

[0208] As used herein, the term "antibody" is used in its broadest sense to include polyclonal and monoclonal antibodies, as well as antigen binding fragments of such antibodies. An antibody useful in a method of the invention, or an antigen binding fragment thereof, is characterized, for example, by having specific binding activity for an epitope of a MICAL or MICAL-Like polypeptide, or a Plexin. In addition, as discussed above, an antibody of the invention can be an antibody that specifically binds a peptide portion of a MICAL, a MICAL-Like polypeptide, or a plexin, particularly a plexin-interacting region of a MICAL or a MICAL-binding region of a plexin.

[0209] The term "binds specifically" or "specific binding activity," when used in reference to an antibody means that an interaction of the antibody and a particular epitope has a dissociation constant of at least about 1×10^{-6}, generally at least about 1×10^{-7}, usually at least about 1×10^{-8}, and particularly at least about 1×10^{-9} or 1×10^{-10} or less. As such, Fab, $F(ab')_2$, Fd and Fv fragments of an antibody that retain specific binding activity for an epitope of a MICAL or MICAL-Like polypeptide, are included within the definition of an antibody.

[0210] The term "antibody" as used herein includes naturally occurring antibodies as well as non-naturally occurring antibodies, including, for example, single chain antibodies, chimeric, bifunctional, human, and humanized antibodies, intrabodies (i.e. intracellularly expressed antibodies, see e.g., Chen, S. Y., et al., *Hum. Gene Ther.* 5:595-601 (1994)), as well as antigen-binding fragments thereof. Such non-naturally occurring antibodies can be constructed using solid phase peptide synthesis, can be produced recombinantly or can be obtained, for example, by screening combinatorial libraries consisting of variable heavy

chains and variable light chains (see Huse et al., *Science* 246:1275-1281 (1989), which is incorporated herein by reference). These and other methods of making, for example, chimeric, humanized, CDR-grafted, single chain, and bifunctional antibodies are well known to those skilled in the art (Winter and Harris, *Immunol. Today* 14:243-246, 1993; Ward et al., *Nature* 341:544-546, 1989; Harlow and Lane, *Antibodies: A laboratory manual* (Cold Spring Harbor Laboratory Press, 1988); Hilyard et al., *Protein Engineering: A practical approach* (IRL Press 1992); Borrabeck, *Antibody Engineering*, 2d ed. (Oxford University Press 1995); each of which is incorporated herein by reference).

[0211] Antibodies that bind specifically with a MICAL or MICAL-Like polypeptide can be raised using the MICAL or MICAL-Like polypeptide, or a fragment thereof, as an immunogen. Where such a polypeptide or fragment thereof is non-immunogenic, it can be made immunogenic by coupling the hapten to a carrier molecule such as bovine serum albumin (BSA) or keyhole limpet hemocyanin (KLH), or by expressing the peptide portion as a fusion protein. Various other carrier molecules and methods for coupling a hapten to a carrier molecule are well known in the art (see, for example, by Harlow and Lane, *supra*, 1988).

[0212] If desired, a kit incorporating an antibody or other agent useful in a method of the invention can be prepared. Such a kit can contain, in addition to the agent, a pharmaceutical composition in which the agent can be reconstituted for administration to a subject. The kit also can contain, for example, reagents for detecting the antibody, or for detecting specific binding of the antibody to a MICAL or MICAL-Like polypeptide. Such detectable reagents useful for labeling or otherwise identifying the antibody are described herein and known in the art.

[0213] Methods for raising polyclonal antibodies, for example, in a rabbit, goat, mouse or other mammal, are well known in the art (see, for example, Green et al., "Production of Polyclonal Antisera," in *Immunochemical Protocols* (Manson, ed., Humana Press 1992), pages 1-5; Coligan et al., "Production of Polyclonal Antisera in Rabbits, Rats, Mice and Hamsters," in *Curr. Protocols Immunol.* (1992), section 2.4.1; each or which is incorporated herein by reference). In addition, monoclonal antibodies can be obtained using methods that are well known and routine in the art (Harlow and Lane, *supra*, 1988). For example,

spleen cells from a mouse immunized with a MICAL or MICAL-Like polypeptide, or an epitopic fragment thereof, can be fused to an appropriate myeloma cell line such as SP/02 myeloma cells to produce hybridoma cells. Cloned hybridoma cell lines can be screened using labeled antigen to identify clones that secrete monoclonal antibodies having the appropriate specificity, and hybridomas expressing antibodies having a desirable specificity and affinity can be isolated and utilized as a continuous source of the antibodies. The antibodies can be further screened for the inability to bind specifically with the MICAL or MICAL-Like polypeptide. Such antibodies are useful, for example, for preparing standardized kits for clinical use. A recombinant phage that expresses, for example, a single chain anti- MICAL or MICAL-Like polypeptide antibody also provides an antibody that can used for preparing standardized kits.

[0214] Methods of preparing monoclonal antibodies well known (see, for example, Kohler and Milstein, *Nature* 256:495, 1975, which is incorporated herein by reference; see, also, Coligan et al., *supra*, 1992, see sections 2.5.1-2.6.7; Harlow and Lane, *supra*, 1988). Briefly, monoclonal antibodies can be obtained by injecting mice with a composition comprising an antigen, verifying the presence of antibody production by removing a serum sample, removing the spleen to obtain B lymphocytes, fusing the B lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones that produce antibodies to the antigen, and isolating the antibodies from the hybridoma cultures.

[0215] Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well established techniques, including, for example, affinity chromatography with Protein-A SEPHAROSE, size exclusion chromatography, and ion exchange chromatography (Coligan et al., *supra*, 1992, see sections 2.7.1-2.7.12 and sections 2.9.1-2.9.3; see, also, Barnes et al., "Purification of Immunoglobulin G (IgG)," in *Meth. :Molec. Biol.* 10:79-104 (Humana Press 1992), which is incorporated herein by reference). Methods of *in vitro* and *in vivo* multiplication of monoclonal antibodies is well known to those skilled in the art. Multiplication *in vitro* can be carried out in suitable culture media such as Dulbecco's Modified Eagle Medium or RPMI 1640 medium, optionally replenished by a mammalian serum such as fetal calf serum or trace elements and growth sustaining

supplements such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages. Production *in vitro* provides relatively pure antibody preparations and allows scale-up to yield large amounts of the desired antibodies. Large scale hybridoma cultivation can be carried out by homogenous suspension culture in an airlift reactor, in a continuous stirrer reactor, or in immobilized or entrapped cell culture. Multiplication *in vivo* can be carried out by injecting cell clones into mammals histocompatible with the parent cells, for example, syngeneic mice, to cause growth of antibody producing tumors. Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection. After one to three weeks, the desired monoclonal antibody is recovered from the body fluid of the animal.

[0216] Therapeutic applications for antibodies disclosed herein are also part of the present invention. For example, antibodies of the present invention can also be derived from subhuman primate antibody. General techniques for raising therapeutically useful antibodies in baboons can be found, for example, in Goldenberg et al., International Patent Publication WO 91/11465 (1991); and Losman et al., *Int. J. Cancer* 46:310, 1990, each of which is incorporated herein by reference. Accordingly, the present invention provides antibodies conjugated to a therapeutic moiety. For example, in certain aspects the present invention provides an anti-MICAL antibody conjugated to a monooxygenase inhibitor, for example EGCG.

[0217] A therapeutically useful anti- MICAL or MICAL-Like polypeptide antibody also can be derived from a "humanized" monoclonal antibody. Humanized monoclonal antibodies are produced by transferring mouse complementarity determining regions from heavy and light variable chains of the mouse immunoglobulin into a human variable domain, and then substituting human residues in the framework regions of the murine counterparts. The use of antibody components derived from humanized monoclonal antibodies obviates potential problems associated with the immunogenicity of murine constant regions. General techniques for cloning murine immunoglobulin variable domains are known (see, for example, Orlandi et al., *Proc. Natl. Acad. Sci., USA* 86:3833, 1989, which is hereby incorporated in its entirety by reference). Techniques for producing humanized monoclonal antibodies also are known (see, for example, Jones et al., *Nature*

321:522, 1986; Riechmann et al., *Nature* 332:323, 1988; Verhoeyen et al., *Science* 239:1534, 1988; Carter et al., *Proc. Natl. Acad. Sci., USA* 89:4285, 1992; Sandhu, *Crit. Rev. Biotechnol.* 12:437, 1992; and Singer et al., *J. Immunol.* 150:2844, 1993; each of which is incorporated herein by reference).

[0218] Antibodies of the invention also can be derived from human antibody fragments isolated from a combinatorial immunoglobulin library (see, for example, Barbas et al., *METHODS: A Companion to Methods in Immunology* 2:119, 1991; Winter et al., *Ann. Rev. Immunol.* 12:433, 1994; each of which is incorporated herein by reference). Cloning and expression vectors that are useful for producing a human immunoglobulin phage library can be obtained, for example, from STRATAGENE Cloning Systems (La Jolla, CA).

[0219] An antibody of the invention also can be derived from a human monoclonal antibody. Such antibodies are obtained from transgenic mice that have been "engineered" to produce specific human antibodies in response to antigenic challenge. In this technique, elements of the human heavy and light chain loci are introduced into strains of mice derived from embryonic stem cell lines that contain targeted disruptions of the endogenous heavy and light chain loci. The transgenic mice can synthesize human antibodies specific for human antigens, and the mice can be used to produce human antibody-secreting hybridomas. Methods for obtaining human antibodies from transgenic mice are described, for example, by Green et al., *Nature Genet.* 7:13, 1994; Lonberg et al., *Nature* 368:856, 1994; and Taylor et al., *Int. Immunol.* 6:579, 1994; each of which is incorporated herein by reference.

[0220] Antibody fragments of the present invention can be prepared by proteolytic hydrolysis of the antibody or by expression in *E. coli* of DNA encoding the fragment. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted $F(ab')_2$. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulfhydryl groups resulting from cleavage of disulfide linkages, to produce 3.5S Fab' monovalent fragments. Alternatively, an enzymatic cleavage using pepsin produces two

monovalent Fab' fragments and an Fc fragment directly (see, for example, Goldenberg, U.S. Patent No. 4,036,945 and U.S. Patent No. 4,331,647, each of which is incorporated by reference, and references contained therein; Nisonhoff et al., *Arch. Biochem. Biophys.* 89:230. 1960; Porter, *Biochem. J.* 73:119, 1959; Edelman et al., *Meth. Enzymol.*, 1:422 (Academic Press 1967), each of which is incorporated herein by reference; see, also, Coligan et al., *supra*, 1992, see sections 2.8.1-2.8.10 and 2.10.1-2.10.4).

[0221] Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light/heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques can also be used, provided the fragments specifically bind to the antigen that is recognized by the intact antibody. For example, Fv fragments comprise an association of V_H and V_L chains. This association can be noncovalent (Inbar et al., *Proc. Natl. Acad. Sci., USA* 69:2659, 1972). Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde (Sandhu, *supra*, 1992). Preferably, the Fv fragments comprise V_H and V_L chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the V_H and V_L domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced into a host cell such as *E. coli.* The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing sFvs are described, for example, by Whitlow et al., Methods: A Companion to Methods in Enzymology 2:97, 1991; Bird et al., *Science* 242:423-426, 1988; Ladner et al., U.S. patent No. 4,946,778; Pack et al., *Bio/Technology* 11:1271-1277, 1993; each of which is incorporated herein by reference; see, also Sandhu, *supra*, 1992.

[0222] Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells (see, for example, Larrick et al.,

Methods: A Companion to Methods in Enzymology 2:106, 1991, which is incorporated herein in its entirety by reference).

[0223] An intrabody comprises at least a portion of an antibody that is capable of immunospecifically binding an antigen and preferably does not contain sequences coding for its secretion. Such antibodies will bind antigen intracellularly. In one embodiment, the intrabody comprises a single-chain Fv ("sFv"). sFvs are antibody fragments comprising the VH and VL domains of antibody, wherein these domains are present in a single polypeptide chain. Generally, the sFv polypeptide further comprises a polypeptide linker between the VH and VL domains which enables the sFv to form the desired structure for antigen binding. For a review of sFvs see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenburg and Moore eds. Springer-Verlag, New York, pp. 269-315 (1994). In a further embodiment, the intrabody preferably does not encode an operable secretory sequence and thus remains within the cell (see generally Marasco, WA, 1998, "Intrabodies: Basic Research and Clinical Gene Therapy Applications" Springer:New York). Generation of intrabodies is well-known to the skilled artisan and is described, for example, in U.S. Patent Nos. 6,004,940; 6,072,036; 5,965,371, which are incorporated by reference in their entireties herein. Further, the construction of intrabodies is discussed in Ohage and Steipe, 1999, *J. Mol. Biol.* 291:1119-1128; Ohage et al., 1999, *J. Mol. Biol.* 291:1129-1134; and Wirtz and Steipe, 1999, *Protein Science* 8:2245-2250, which references are incorporated herein by reference in their entireties.

[0224] In another embodiment, the present invention provides a double-stranded RNA molecule that includes a first RNA strand that specifically hybridizes to an mRNA encoding a MICAL polypeptide. The double-stranded RNA molecule also includes a second RNA strand that is the reverse complement of the first strand. The double-stranded molecule is at least 15 base pairs in length. Double stranded RNA is involved in RNA interference, the process by which double-stranded RNA induces the silencing of homologous endogenous genes. (Hammond, S. M., et al., *Nat. Rev. Genet.* 2(2):110-9 (2001)). Accordingly, double stranded RNA of this embodiment inhibit expression of a MICAL polypeptide, thereby promoting axonal growth and target formation.

[0225] In another embodiment the present invention provides a method for identifying an agent that affects an activity of a MICAL or MICAL-Like protein. As such, the present invention provides screening methods for agents that affect MICAL protein or MICAL-Like protein activity. The method typically includes contacting a MICAL polypeptide, or a functional portion thereof or a MICAL-Like polypeptide, or a functional portion thereof, or a cell expressing at least one of these polypeptides, with a candidate agent, and determining whether the agent affects an activity of the polypeptide. The activity of the MICAL protein, can be any of the activities identified herein for a MICAL polypeptide.

[0226] For example, a method according to this embodiment can identify an agent that affects MICAL monooxygenase activity or plexin interacting activity. Methods for identifying monooxygenase activity and plexin interacting activity are known in the art. Examples of these methods are provided herein. For example, an immunoprecipitation experiment can be performed in the presence of plexA and a MICAL polypeptide, wherein the plexA and/or MICAL polypeptide are contacted with an on-test agent. It can then be determined whether the agent affected binding of PlexA and the MICAL polypeptide. Agents that affect binding are candidate agents for treatment of disorders such as spinal cord injury, since they are expected to inhibit semaphorin-mediated axonal repulsion.

[0227] In one aspect of this embodiment, the method can identify an agent that affects a semaphorin-mediated process. That is, the activity of a MICAL protein can be participation, typically regulation, of a semaphorin mediated process. For example, a semaphorin-mediated process can be semaphorin-mediated axonal repulsion. In particular embodiments, the semaphorin-mediated process is mediated by semaphorins 1A and 3A, 4A, or a class 7 semaphorin.

[0228] In aspects of this embodiment that include a cell, the cell can be virtually any cell. Recombinant cells can be produced using standard techniques as disclosed herein, that express a MICAL polypeptide using polynucleotides and vectors of the present invention. In embodiments where the method identifies an agent that affects a MICAL activity and/or a semaphorin-mediated process, the cell, for example, can be a cell of a type that is known to include a semaphorin-mediated process, as discussed hereinbelow, such as a cell of the

immune system, for example a B cell or a T cell, a cell of neuronal origin, a cell with a transformed phenotype, a cardiac cell, or a neural crest precursor cell of a cardiac cell.

[0229] In another aspect of this embodiment, the present invention provides a method for identifying an agent that affects axonal guidance regulatory activity, monooxygenase activity, actin binding activity and/or plexin-interacting activity. The method typically includes contacting an isolated polypeptide of the present invention that has axonal guidance regulatory activity, monooxygenase activity, actin binding activity, and/or plexin-interacting activity, or a cell expressing the polypeptide, with a candidate agent. Next, axonal guidance regulatory activity, monooxygenase activity, actin binding activity, and/or plexin-interacting activity is compared in the presence versus absence of the agent. A difference in activity is indicative of an agent that affects axonal guidance regulatory activity, monooxygenase activity, and/or plexin-interacting activity. In this aspect of the invention, the cell is typically a cell of neuronal original, such as a neuron that recombinantly expresses the MICAL.

[0230] Methods for determining axonal guidance regulatory activity (i.e. axon guidance regulatory activity assays), for example semaphorin axonal repulsion activity, are known in the art. Some of these methods are disclosed herein. For example, an *in vitro* method such as a rat DRG growth cone repulsion assay using Sema 3A-secreting 293 cells, as disclosed above, can be employed (Figure 4A; Messersmith et al., 1995). It will be recognized however, that virtually any assay of semaphorin mediated axon repulsion can be employed in the methods of the present invention for identifying an agent that affects axonal guidance regulatory activity. Methods for assessing plexin-interacting activity and monooxygenase activity are provided herein.

[0231] In another embodiment, the present invention provides a method for screening for an agent that modulates an activity of a MICAL or MICAL-Like polypeptide. The method includes contacting a cell that recombinantly expresses a MICAL or MICAL-Like polypeptide with a candidate agent. Then a phenotypic or physiological trait of the cell is compared in the presence or absence of the candidate agent. A difference in the phenotypic or physiological trait indicates that the agent modulates the activity of the MICAL polypeptide. The phenotypic or physiological trait can involve dynamics of the

cytoskeleton, or can be axon guidance, cell proliferation or invasiveness, or an immune response.

[0232] In another embodiment, the present invention provides a method for screening for an agent that modulates the expression of a MICAL or MICAL-Like polypeptide. The method includes contacting a cell with a candidate agent. Then comparing the expression of the MICAL or MICAL-Like polypeptide, for example in the presence or absence of the candidate agent. A difference in the expression indicates that the agent modulates the expression of the MICAL polypeptide. In one aspect, the level of mRNA encoding MICAL is compared. In another aspect, the level of the MICAL polypeptide is compared.

[0233] Specificity can be further analyzed, for example, by assaying for an activity or phenotype in cells that recombinantly overexpress MICAL and comparable cells that do not recombinantly overexpress MICAL, or cells in which MICAL has been knocked out or reduced versus comparable cells in which MICAL is expressed at normal levels.

[0234] As illustrated in the Examples herein, MICALs are susceptible to small molecule inhibitors that affect their ability to oxidize their substrate. For example, gallic acid derivatives, including the green tea component (-)-epigallocatechin gallate (EGCG), appear to be potent and selective inhibitors of MICALs. It will be recognized that based on the identification of these small molecule inhibitors of semaphorin-mediated axonal repulsion, it will be recognized that other small molecule inhibitors can be identified.

[0235] All available evidence points to the plexin cytoplasmic domain as an essential signal transducing domain for signaling class 3 semaphorin repulsion (Cheng et al., 2001; Takahashi and Strittmatter, 2001). Sema3A appears to utilize neuropilin-1 in combination with A class plexins to signal repulsive axon guidance. As illustrated in the Examples herein, agents that affect MICAL axonal guidance regulatory activity can be identified *in vitro*. For example, a rat DRG growth cone repulsion assay using Sema 3A-secreting 293 cells, as disclosed above, can be employed (Figure 4A; Messersmith et al., 1995). NGF-dependent DRG axons exhibit little to no outgrowth towards Sema3A-secreting 293 cell aggregates (Figure 4C). However, this repulsion can be neutralized in a specific and dose-dependent manner by inclusion of an inhibitor such as EGCG or EC, in the growth media

(Figure 4C). As illustrated in the Example, like EGCG, EC is capable of completely neutralizing Sema-3A-dependent repulsion in a dose-dependent manner, but a much higher EC concentration is required (Figures 4C).

[0236] As will be understood, typically a method according to this embodiment of the invention includes a control sample wherein the polypeptide or cell is not contacted with the agent. However, known control values or qualitative results can also be used, such that a control sample does not need to be included each time the method is performed. For example, a visual microscopic comparison of axonal outgrowth can be performed using methods of the present invention that utilize the DRG rat growth cone repulsion assay. In addition, results of the rat DRG growth cone repulsion assay can be quantitated as illustrated in Fig. 4C. For example, results can be scored as the ratio of the axon lengths on the proximal and distal sides of the explant (P/D ratio), and on Sema3A-mediated growth cone collapse indicated as % collapsed growth cones. Therefore, known values for controls can be calculated and used to establish a baseline using known statistical methods, above which significant inhibition of the repulsion is established, thereby identifying an agent that affects axon guidance regulatory activity, most particular semaphorin-mediated axon repulsion.

[0237] The agent can affect axonal guidance regulatory activity, monooxygenase activity, and/or plexin interacting activity by enhancing or inhibiting this activity. The agent can be a small molecule, such as an antioxidant flavonoid, an antisense polynucleotide, a MICAL-like polypeptide or fragment thereof, a mutant MICAL polypeptide, a mutant MICAL-Like polypeptide, an anti-MICAL antibody, a double stranded RNA, or a peptidomimetic.

[0238] As indicated herein, anti-oxidant flavonoids such as ECGC and EC are inhibitors of semaphorin-mediated axonal repulsion. Therefore, particularly important candidate agents include those with anti-oxidant activity, especially anti-oxidant flavonoids or other green tea polyphenols. A variety of antioxidant flavonoids have been identified and can be analyzed to determine whether they affect semaphorin-mediated axonal repulsion. For example, the antioxidant flavonoid can be a gallic acid derivative or another flavoprotein monooxygenase inhibitor. Examples of gallic acid derivatives that can be analyzed using

methods of the present invention include, but are not limited to, (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3''-*O*-methyl-EGCG, 3''-*O*-methyl-ECG, 3''-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, and *n*-cetyl gallate.

[0239] Additionally, it will be recognized that new antioxidant flavonoids, for example new gallic acid derivatives can be synthesized and tested using the methods of this aspect of the present invention for the ability to affect axonal guidance regulatory activity. Furthermore, flavoprotein monooxygenase inhibitors, which include but are not limited to certain anti-oxidant flavonoids can be analyzed for an affect on axonal guidance regulatory activity using methods of this embodiment of the present invention.

[0240] An agent tested by the screening embodiments of the present invention can include a known oxidase inhibitor. For example, an oxidase inhibitor outlined in Cross, *Free Radical Biology and Med.*, 8:71-93 (1990) including, for example, DPI, ibuprofen, and aspirin. Furthermore, the screening assay can be used to test other agents including inhibitors of similar flavoprotein monooxygenase family members (See e.g., Cross (supra.); and Arnould and Camadro, *Proc. Nat. Acad. Sci.*, 95:10553-10558 (1998)).

[0241] In certain embodiments, a screening method of the invention can be performed, for example, by contacting under suitable conditions a MICAL, or a functional peptide portion thereof, a plexin such as plexin A, and an agent to be tested. The MICAL, the plexin and the agent can be contacted in any order as desired. As such, the screening method can be used to identify agents that can competitively or non-competitively inhibit MICAL binding to plexin, agents that can mediate or enhance MICAL binding to the plexin, agents that can induce dissociation of specifically bound MICAL from the plexin, and agents that otherwise affect the ability of MICAL to regulate axon guidance, such agents having agonist or antagonist activity. Appropriate control reactions are performed to confirm that the action of the agent is specific with respect to the MICAL.

[0242] Suitable conditions for performing a screening method of the invention can be any conditions that allow MICAL to specifically interact with plexin, that provide a

semaphorin-plexin repulsive axon guidance activity, and/or that support MICAL monooxygenase activity, including methods as disclosed herein or otherwise known in the art. Thus, suitable conditions for performing the screening assay can be, for example, *in vitro* conditions using a substantially purified MICAL and/or plexin; cell culture conditions, utilizing a cell that normally expresses a semaphorin-plexin A mediated repulsive axon activity and a MICAL polypeptide, for example, a neuron, or a cell that has been genetically modified to express a semaphorin-plexin A mediated repulsive axon activity including expression of a MICAL polypeptide; or *in situ* conditions as occur in an organism.

[0243] A screening method of the invention also can be performed using the methods of molecular modeling. The utilization of a molecular modeling method provides a convenient, cost effective means to identify those agents, among a large population such as a combinatorial library of potential agents, that are most likely to interact specifically with MICAL or a plexin, thereby reducing the number of potential agents that need to be screened using a biological assay. Upon identifying agents that interact specifically with a MICAL or a plexin, for example Plexin A, using a molecular modeling method, the selected agents can be examined for the ability to modulate an effect of a MICAL on a cell, such as regulation of axon guidance, using the methods disclosed herein.

[0244] The ability of a test agent to modulate an effect of MICAL can be detected using methods as disclosed herein or otherwise known in the art. The term "test agent" or "test molecule" is used broadly herein to mean any agent that is being examined for agonist or antagonist activity in a method of the invention. Although the method generally is used as a screening assay to identify previously unknown molecules that can act as agonist or antagonist agents as described herein, the methods also can be used to confirm that a agent known to have a particular activity in fact has the activity, for example, in standardizing the activity of the agent.

[0245] Further assays for testing the specificity of a candidate agent for MICAL, for example, test for lack of inhibition of steroid 5 alpha-reductase, NADPH-cytochrome P450 reductase, telomerase, MMP-2, MMP-9, and phenol sulfotransferase. These enzymes are known to be inhibited by EGCG in addition to its inhibition of flavoprotein

monooxygenases. Therefore, by using these assays, agents can be identified that are more specific inhibitors of MICALs than EGCG.

[0246] A screening method of the invention provides the advantage that it can be adapted to high throughput analysis and, therefore, can be used to screen combinatorial libraries of test agents in order to identify those agents that can modulate an effect of MICAL on a cell, including those agents that can alter a specific interaction of MICAL and a plexin, and those that otherwise modulate MICAL axon guidance regulatory activity. Methods for preparing a combinatorial library of molecules that can be tested for a desired activity are well known in the art and include, for example, methods of making a phage display library of peptides, which can be constrained peptides (see, for example, U.S. Patent No. 5,622,699; U.S. Patent No. 5,206,347; Scott and Smith, *Science* 249:386-390, 1992; Markland et al., *Gene* 109:13-19, 1991; each of which is incorporated herein by reference); a peptide library (U.S. Patent No. 5,264,563, which is incorporated herein by reference); a peptidomimetic library (Blondelle et al., *Trends Anal. Chem.* 14:83-92, 1995; a nucleic acid library (O'Connell et al., *supra*, 1996; Tuerk and Gold, *supra*, 1990; Gold et al., *supra*, 1995; each of which is incorporated herein by reference); an oligosaccharide library (York et al., *Carb. Res.*, 285:99-128, 1996; Liang et al., *Science*, 274:1520-1522, 1996; Ding et al., *Adv. Expt. Med. Biol.*, 376:261-269, 1995; each of which is incorporated herein by reference); a lipoprotein library (de Kruif et al., *FEBS Lett.*, 399:232-236, 1996, which is incorporated herein by reference); a glycoprotein or glycolipid library (Karaoglu et al., *J. Cell Biol.*, 130:567-577, 1995, which is incorporated herein by reference); or a chemical library containing, for example, drugs or other pharmaceutical agents (Gordon et al., *J. Med. Chem.*, 37:1385-1401, 1994; Ecker and Crooke, *Bio/Technology*, 13:351-360, 1995; each of which is incorporated herein by reference). Polynucleotides can be particularly useful as agents that can modulate a specific interaction of MICAL and a plexin because nucleic acid molecules having binding specificity for cellular targets, including cellular polypeptides, exist naturally, and because synthetic molecules having such specificity can be readily prepared and identified (see, for example, U.S. Patent No. 5,750,342, which is incorporated herein by reference).

[0247] In view of the present disclosure, it will be recognized that various animal model systems can be used as research tools to identify agents useful for practicing a method of the invention. For example, as discussed above, transgenic flies, mice or other experimental animals can be prepared and the transgenic non-human organism can be examined directly to determine the effect produced by expressing various levels of a particular agent in the organism.

[0248] As disclosed herein, MICALs can exert their activity, at least in part, through a Semaphorin-Plexin mediated pathway which can be associated with various pathological conditions. As such, the present invention provides new targets for the treatment of various conditions, especially regeneration of damaged neurological tissue, such as damage resulting from spinal cord injury, abnormal immune cell function, and abnormal cell motility such as cancer cell motility. Accordingly, the present invention provides methods for ameliorating the severity of a pathological condition in a subject, wherein the pathologic condition is characterized, for example, by an inability of neurons to regenerate proper axonal connections.

[0249] In another embodiment, the present invention provides a method for affecting axonal guidance regulatory activity. The method includes contacting a cell, typically a neuron, that expresses a polypeptide of the invention such as a MICAL polypeptide, with an agent that affects axonal guidance regulatory activity or monooxygenase activity. Typically, the method is performed *in vivo* and includes inhibiting axonal guidance regulatory activity by contacting the cell with an antioxidant. The cell can be contacted chronically with the antioxidant. The axonal guidance activity is typically semaphorin-mediated axonal repulsion. As such, in another embodiment, the present invention provides a method for affecting a semaphorin-mediated process by contacting a cell that expresses a MICAL polypeptide of the invention with an effective amount of an agent that affects axonal guidance regulatory activity. The agent can be, for example, a small molecule, a polypeptide or fragment thereof, a peptidomimetic, or an antisense polynucleotide, as discussed herein.

[0250] Not to be limited by theory, as disclosed herein MICALs regulate semaphorin-mediated axonal repulsion through a mechanism that requires their monooxygenase activity.

Therefore, as illustrated in the Examples section, antioxidants such as flavonoids inhibit semaphorin-mediated axonal repulsion, apparently by overcoming the effects of MICAL monooxygenase activity.

[0251] As mentioned above, following spinal cord injury in humans, axons fail to reestablish their connections, which results in paralysis and loss of sensation of the affected area. During development, inhibition of axon growth plays a role in forming the nervous system. Axons are guided to their targets by molecules that attract them as well as by those that inhibit (i.e., repel) them. The inhibition of semaphorin-mediated axonal repulsion can allow axons to reach new targets. Therefore, methods for treating spinal cord injuries can focus on inhibiting axonal repulsion (See e.g., Schwab, M. E., *Science*, 295:1029 (2002); and Fournier, A. E., and Strittmatter, S. M., *Current Opinion in Neurobiol.*, 11:89 (2001)). Accordingly, methods of this embodiment of the invention are useful for example, for inducing regeneration of axons after spinal cord injury.

[0252] In another embodiment, the present invention provides a method for treating a neurological condition in a subject, that includes contacting in the subject, a cell of the central nervous system and/or peripheral nervous system having a disrupted axonal connection, or a cell that affects axonal growth of the central nervous system and/or peripheral nervous system cell, or surrounding tissue, with an amount of an agent that modulates the activity or expression of a MICAL polypeptide, the amount being effect to modulate axonal guidance regulatory activity, axon out-growth, monooxygenase activity, or plexin-interacting activity. The subject in certain aspects is a human patient in need of the treatment. The neurological condition is any neurological condition in which a treatment strategy includes promoting axon growth. For example, neurological conditions treatable using methods of the present invention include a spinal cord injury, traumatic brain injury, neuropathic pain, Parkinson's Disease, Amyotrophic Lateral Sclerosis (ALS), ischemic injury, Alzheimer's Disease, Multiple Sclerosis, Huntington' s chorea, multiple system atrophy, progressive supranuclear palsy, traumatic brain injury, neuropathic pain, ischemic injury, a neuropathy resulting from a stroke, a peripheral neuropathy resulting from chemotherapy, or a peripheral neuropathy resulting from diabetes. The neurodegenerative disease may be associated with a bacterial, viral or other infection, such as damage caused

by HIV or herpes viral infections, encephalitis, and Creutzfeldt-Jacob disease and kuru or may be due to the effects of a drug or toxin. Surrounding tissue is any substrate through which an axon need to re-grow or re-establish its connections.

[0253] In certain aspects of this embodiment of the invention, the agent contacts a site in need of axonal growth or regrowth chronically. For example, the agent is applied for a length of time sufficient to promote neurorestoration. In general, the length of time of chronic application of an agent for the present invention is longer than the time period required to protect injured neurons (i.e. neuroprotection) from the harmful cascade of secondary events that follow injury, for example detrimental inflammatory responses and death of neurons and glia (i.e. secondary death due to reactive oxidative species (lipid peroxidation)). The length of time for chronic agent contact in the present invention is the time necessary to continue to stimulate axonal growth, to guide axons to their targets, and/or to establish new functional synapses. In certain aspects of the present invention, an agent can be applied initially to save neurons (neuroprotection), and continued over longer periods to promote neurorestoration. Therefore, the present invention in certain aspects, couples neuroprotection and neurorestoration through delivery of agents a short time after neurological damage, followed by long-term administration of the agent.

[0254] In one aspect of the present invention, the agent is applied for at least 1, 2, 7, or 14 days, or 1, 2, 3, 4, 5, 6, 12, 24, 36, 48, 60 months, either continually or repeatedly, for example by use of gene therapy, after identification or suspicion of a neurological condition. Other embodiments of the invention that include beneficial aspects directed at chronic application of an agent include for example, a method for affecting axon growth, a method for affecting a plexin mediated process, a method for treating a neurological disorder involving failure of axon regrowth, and a method for inducing regrowth and preventing inhibition of an injured process of a neuron.

[0255] As indicated herein, monooxygenase inhibitors and anti-oxidant flavonoids, including those that are monooxygenase inhibitors such as ECGC and EC, are inhibitors of semaphorin-mediated axonal repulsion. Accordingly, monooxygenase inhibitors and anti-oxidant flavonoids can be used as the agent in methods of various embodiments of the present invention, to inhibit axonal guidance regulatory activity. These embodiments of the

invention include methods for treating a neurological condition, methods for affecting axonal guidance regulatory activity, methods for affecting axon growth, methods for affecting a plexin-mediated process, methods for treating a neurological disorder involving a failure of axon regrowth, and methods for inducing regrowth of an injured process of a neuron.

[0256] As mentioned above, a variety of monooxygenase inhibitors and anti-oxidant flavonoids have been identified and can be used to inhibit axonal guidance regulatory activity, such as semaphorin-mediated axonal repulsion. The anti-oxidant flavonoid can be a gallic acid derivative such as ECGC or EC, another flavoprotein monooxygenase inhibitor, another green tea polyphenol. Examples of gallic acid derivatives that can be used to affect axonal guidance regulatory activity in embodiments of the present invention include, but are not limited to, ECGC, EC, (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3"-*O*-methyl-EGCG, 3"-*O*-methyl-ECG, 3"-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, and *n*-cetyl gallate.

[0257] An agent used in the methods of the present invention for affecting axonal guidance regulatory activity or monooxygenase activity, in certain aspects, is a known oxidase inhibitor. For example, an oxidase inhibitor reported in Cross, *Free Radical Biology and Med.*, 8:71-93 (1990) including DPI, ibuprofen, and/or aspirin. Furthermore, the screening assay can be used to test other agents including inhibitors of similar flavoprotein monooxygenase family members (See e.g., Cross (supra.); and Arnould and Camadro, *Proc. Nat. Acad. Sci.*, 95:10553-10558 (1998)).

[0258] Various embodiments of the present invention can further include contacting the cell with an agent that modulates MICAL activity and affects axon regeneration. The embodiments include methods for treating a neurological condition, methods for treating a neurological disorder involving a failure of axon regrowth, and methods for inducing regrowth of an injured process of a neuron, as described herein. In certain aspects, the agent that affects axon regeneration promotes axon regeneration. Accordingly, the agent can be a

neurotrophic factor, a mechanical bridge, and/or a stem cell (see e.g., Blesch A., et al., *Brain Res. Bulletin*, 57:83, 2002; and Cao, Q., et al., *J. Neurosci. Res.*, 68: 501 (2002)).

[0259] Mechanical bridges include, for example, genetically engineered cells, stem cells, fetal tissue, Schwann cells, olfactory ensheathing glia, as discussed below. Neurotrophins are molecules with closely related structures that are known to support the survival of different classes of embryonic neurons. Virtually any neurotrophin can be used in methods of the present invention, including, for example, nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin 3 (NT 3), neutortrophin-4/5 (NT-4/5), glia cell line-derived neurotrophic factor (GDNF), and leukemia inhibitory factor (LIF).

[0260] Neurotrophins can be delivered to a site of severed axons, for example, by use of gene therapy. For example, viral vectors, including retroviral vectors, as described herein, capable of infecting neurons or glia can provide a localized source of trophic factors, such as neurotrophins to stimulate axonal outgrowth. Genetically modified cells grafted to a lesion site in the spinal cord can provide not only augmented amounts of trophic molecules at the injury site, but also a potential axonal growth substrate. Thus, genetically modified cells can provide mechanical bridges for growing axons to potentially connect injured spinal cord regions(Blesch et al., 2002). Furthermore, genetically modified cells can be used for long-term delivery of a neurotrophin, in a regulated manner. Furthermore, gene therapy can be used to deliver MICAL inhibitors, such as mutant MICALs or MICAL-Like proteins, chronically to a site of spinal cord injury.

[0261] Cell types used for grafting in gene therapy, can include for example fibroblasts, Schwann cells, and neural stem cells (see e.g., Brecknell J. E., et al., *Neuroscience*, 74(3):775-84 (1996)). Furthermore, autologous cells can be used to avoid immune responses and graft rejection. Gene therapy for use in the present invention can include *in vivo* or *ex vivo* gene therapy.

[0262] Methods of the present invention that include providing a stem cell as well as an anti-oxidant typically involve providing both the stem cell and the anti-oxidant to a site of neuronal damage, such as a site of spinal cord injury. For example, the stem cell can be grafted to a site of spinal cord injury.

[0263] A variety of stem cells can be used with the methods of the present invention. The stem cells are typically neural stems cells (i.e., stem cells that give rise to cells of neuronal lineage). The stem cell can be an embryonic stem cell or an adult stem cell and can be isolated, for example from blood or bone marrow (See e.g., Kabos, P., et al., *Exp Neurol.*, 178(2):288-93 (2002)). Neural stem cells typically express the marker, nestin. The stem cells are capable of giving rise to neurons and glial cells.

[0264] Neural stem cells could be induced towards neuronal phenotypes to allow the replacement of spinal neurons lost after injury, toward astrocytes to restore the non-neuronal milieu of the pre-injured spinal cord, or towards oligodendroglia to allow remyelination. Either neuronal or glial lineages could be useful for reconstituting a permissive substrate for regenerating axons to extend over sites of injury (Blesch et al., 2002). The cells can be modified to produce a recombinant neurotrophic factor.

[0265] Other mechanical bridges of the invention in certain aspects, include an axon outgrowth promoting molecule such as netrin, laminin, collagen, or artificial polymer-based substrates. In another embodiment, the present invention provides a method for affecting axon growth, that includes contacting a neuronal lineage cell with an agent that inhibits axonal guidance regulatory activity of a polypeptide as set forth in SEQ ID NO:2 (human MICAL-1), SEQ ID NO:4 (human MICAL-2), SEQ ID NO:6 (human MICAL-3), or SEQ ID NO: 8 (*Drosophila* MICAL long isoform), SEQ ID NO:10 (*Drosophila* MICAL medium isoform), SEQ ID NO:12 (*Drosophila* MICAL short isoform).

[0266] The neuronal lineage cell can be for example, a neuron or an oligodendrocyte (*J. Neurosci.* Jul 15;22(14):5992-6004 (2002)). Furthermore, the neuronal lineage cell can be a neural stem cell, or a neuronal cell derived from an isolated stem cell. In this aspect, the stem cell can be introduced into a subject at the site of a spinal cord injury, and contacted with the agent before introduction into the subject, and/or at the site of the spinal cord injury. Accordingly, the cell can be contacted with the agent *in vivo* or *in vitro*.

[0267] In another embodiment, the present invention provides a method for affecting a plexin-mediated process or a semaphorin-mediated process, that includes contacting a cell that carries out the plexin-mediate process or semaphorin-mediated process, such as a cell

expressing a MICAL-polypeptide of the present invention, with an effective amount of an agent, for example an antioxidant such as a monooxygenase inhibitor, that affects a MICAL polypeptide activity. The agent affects the plexin-mediated process or the semaphorin-mediated process. For example, where the plexin-mediated process is axonal regrowth, the agent is chronically contacted with the cell.

[0268] For this embodiment of the invention, virtually any cell can be used. For example a recombinant cell that expresses a MICAL polypeptide can be used. The recombinant cell can be obtained using a vector of the present invention and transfection or transformation methods well known in the art. The cell in this embodiment of the invention can be a cell of the nervous system, such as a neuron, an immune cell, a cancer cell, and a cardiac cell, particularly a cardiac neural crest.

[0269] As is indicated herein by the analysis of MICAL proteins in a model invertebrate system (*Drosophila*) and in the vertebrate nervous system, the interactions between all semaphorins, both secreted and transmembrane, with plexin family members are likely to involve essential interactions and functions provided by MICAL proteins. The present disclosure suggests that the N-terminal MICAL monooxygenase domain is essential for semaphorin/plexin-mediated neuronal repulsion. Therefore, MICALs and the subclass of monooxygenase they belong to are targets with regard to neuronal regeneration following injury and various strategies designed to promote neuronal regeneration following neurodegeneration. Additionally, since plexins bind homophilically to other plexins (*Neuron*, 14: 1189-99 (1995)) they may signal in a semaphorin-independent, but MICAL-dependent, manner.

[0270] The cell included in methods of the present invention, including methods for affecting a plexin-mediated or semaphorin-mediated process, in certain aspects is a cell of the immune system. For example, the cell of the immune system can be a lymphocyte, such as a B-cell, a T-cell, or precursor thereof, a monocyte, or a phagocyte. Other cell types for methods of the present invention include cancer cells, particularly metastatic cancer cells, and cardiac cells, particularly cardiac cells from the neural crest. In certain aspects the plexin-mediated process is mediated by Plexin A or Plexin B. In certain aspects the

semaphorin-mediated process is mediated by Sema 1a, Sema 3a, Sema 7a, or a class 4 semaphorin.

[0271] Semaphorins and plexins have been shown to provide important functions in the immune system (See e.g., Ventura, A., and Pelicci, P. G., "Semaphorins: Green Light for Redox Signaling?" *Science's STKE*, pgs. 1-3 (2002)). In addition to what appear to be a plexin-independent function of the class 4 semaphorin Sema4D (i.e. CD100) in the enhancement of B-cell responses via the inactivation of CD72, viral semaphorin interactions with plexin C1 (aka VESPR) induce robust responses in human monocytes (i.e. cell aggregation and the expression of pro-inflammatory cytokines). The mammalian orthologue of one of these viral semaphorins, called Sema7A, may therefore also be involved in immunomodulation events. Further, the interactions between class 4 semaphorins such as CD100 with B class plexins, interactions which are well-characterized in the vertebrate neurons, may function in the immune system. The expression of all class B plexins in the immune system has not been exhaustively determined, and it is likely that neuronal class 4 semaphorin/Plexin B functions have immune system counterparts.

[0272] Finally, the initial characterization of MICAL as a CasL interacting protein was carried out using human thymus cells as a starting source of tissue. Cas family members are important for TCR and b1 integrin-induced immunological reactions in lymphocytes, including interleukin-2 production and various migratory responses. Thus, though at present it is difficult to predict exact roles in immune system function, it is likely that MICALs will be involved in plexin immune system function, that modulation of Cas protein function will follow, and that the disclosure herein with respect to characterization of flavonoids which are likely MICAL antagonists will at the least provide a set of compounds with potent effects on immune system function.

[0273] The plexin-mediated process or the semaphorin-mediated process can be semaphorin 1a-PlexA-mediated repulsive axon guidance. In another aspect, however, the plexin-mediated process or semaphorin-mediated process affected by antioxidants in the methods of the present invention can include cell migration, for example the migration of cancer cells (see e.g., Trusolino, L., and Comoglio, P.M., Nat., *Rev. Cancer,* 2:289 (2002)). Studies by J. Mina and colleagues, for example, have implicated class 3 semaphorins in

certain small cell lung carcinomas. Sema3F, and more recently the closely linked gene for Sema3B, have both been associated with genetic lesions in lung cancer cell lines, and the most recent evidence suggests that Sema3B may be a key determinant in these lung cancers. Given demonstrations of class 3 semaphorin mediated effects on neural crest cell morphology and GABAergic cortical neuronal migration, coupled with the *in vitro* semaphorin collapse assays recently developed in tissue culture cell lines (COS cells, *Cell*, 99:59-69 (1999)), if seems reasonable to think that non-neuronal cell adhesivity and migration are influenced by these secreted repellents. Based upon the present disclosure, which implicate MICAL monooxygenase function in Sema3A repulsion in vertebrates, it is expected that class 3 semaphorin function can be either attenuated or enhanced by using anti-oxidants that affect MICAL function.

[0274] Class 4 semaphorins, which as described are likely to play key roles in the immune system, have also recently been implicated in regulating invasive growth by Comoglio and colleagues. Sema3F in a liver cell line apparently can mediate invasive growth via a coupling of PlexinB1 and Met receptor signaling. This requirement for a plexin and an associated membrane bound co-receptor component for a functional semaphorin response is reminiscent in certain respects of observations in *Drosophila* from Goodman and colleagues showing that Plexin A requires offtrack (OTK), a protein related to receptor tyrosine kinases, for plexin-mediated semaphorin repulsion. This work suggests that Plexin B1, and perhaps many other plexins, can regulate cell migration, either positively or negatively. It is interesting to note that recent work by T. Hunter and colleagues demonstrates a requirement for both FAK and CasL function for Ephrin/Eph receptor mediated attractive responses.

[0275] A role has also been established for plexinA2 and semaphorin signaling in cardiac neural crest cells. For example, studies have shown that PlexinA2 is expressed in migrating and postmigrating cardiac neural crest cells in the mouse (Brown et al., *Development* 128:3071 (2001)). Furthermore, it has been shown that PlexinA2-expressing cardiac neural crest cells are patterned abnormally in several mutant mouse lines with congenital heart disease including those lacking Semaphorin 3C. (*Id.*).

[0276] Additionally, reports have suggested a role of plexins in numerous other diseases. For example, studies have revealed that at least certain semaphorins are dysregulated and/or downregulated in patients with Alzheimer's Disease and Down's Syndrome. (Andorn, A. C., and Kalaria, R.N., *Acta. Neurochir. Suppl.* (Wien) 70:212-5 (1997); Lubec, G., J *Neural Transm Suppl.*, 57:161-77 (1999); and Hirsch et al., *Brain Res.*, 27:67-79 (1999)). Other studies have identified a plexin family member as involved in polycystic kidney and hepatic disease (Onuchic, L.F., et al., *Pediatr. Res.* 52:830 (2002)). Furthermore, single nucleotide polymorphisms (SNPs) related to Rett syndrome have been identified in a Plexin gene (Dahle, A. R., et al., Am. J. Med. Genet. 3:69 (2000)).

[0277] Finally, the very well documented roles of Cas family members in regulating non-neuronal cell morphology and various growth characteristics also suggest understanding that MICALs are linked to CasL could be invaluable for controlling certain cellular behaviors. p130Cas was first identified and a target of hyperphosphorylation by oncogenic Src and Crk. Since then, many *in vitro* experiments have implicated Cas proteins in the maintenance of the transformed cell state, however to date *in vivo* support for Cas function in cell transformation is still lacking. Nevertheless, the demonstrated roles for Cas proteins in cell cycle progression, the regulation of cell shape, and for the induction of cell migration, make this protein an attractive target for controlling a myriad of cellular functions, from osteoclast activation to vasculogenesis and angiogenesis. How MICAL regulates Cas is at present unknown and is a major focus for our own research. However, given well-established interactions between Cas family members and a variety of proteins essential for the establishment and maintenance of cell shape, cell-cell interactions, and morphological changes including migration (including, but not limited to, FAK, RAFTK, Crk, Fyn, Yes, Abl, Grb2, several phosphatases including PTP1B, Nck, and Src), it is likely that MICALs play a role in these processes through Cas interactions.

[0278] A method of the invention for the various embodiments that include contacting a cell or a MICAL with an agent such as an antioxidant, can be performed, for example, by contacting under suitable conditions a target cell and an agent that affects a MICAL function such as MICAL axon guidance regulatory activity. Suitable conditions can be provided by placing the cell, which can be an isolated cell or can be a component of a tissue

or organ, in an appropriate culture medium, or by contacting the cell *in situ* in an organism. For example, a medium containing the cell can be contacted with an agent the affects the ability of a MICAL to specifically interact with a plexin expressed on the cell, or with an agent that affects MIXAL axon guidance regulatory activity in the cell. In general, the cell is a component of a tissue or organ in a subject, in which case contacting the cell can comprise administering the agent to the subject. However, the cell also can be manipulated in culture, then can be maintained in culture, administered to a subject, or used to produce a transgenic nonhuman animal.

[0279] An agent useful in a method of the invention can be any type of molecule, for example, a polynucleotide, a peptide, a peptidomimetic, peptoids such as vinylogous peptoids, a small organic molecule, or the like, and can act in any of various ways to affect a MICAL function such as axon guidance regulatory activity. The agent can act to alter a semaphorin-mediated pathway in the cell. In addition, the agent can be an agonist, which mimics or enhances the effect of MICAL on a cell, for example, the ability of MICAL to specifically interact with a plexin, thereby increasing MICAL axon guidance regulatory activity in the cell; or can be an antagonist, which reduces or inhibits the effect of MICAL on a cell, thereby reducing or inhibiting MICAL axon guidance regulatory activity. For example, administration of an antagonist can result in axon growth and new positioning to allow formation of new targets by inhibiting semaphorin-mediated axonal repulsion activity.

[0280] As used herein, the term "specific interaction" or "specifically binds" or the like means that two molecules form a complex that is relatively stable under physiologic conditions. The term is used herein in reference to various interactions, including, for example, the interaction of MICAL and a plexin such as plexin A, the interaction of the intracellular components of a semaphorin-mediated pathway, and the interaction of an antibody and its antigen. A specific interaction can be characterized by a dissociation constant of at least about 1×10^{-6} M, generally at least about 1×10^{-7} M, usually at least about 1×10^{-8} M, and particularly at least about 1×10^{-9} M or 1×10^{-10} M or greater. A specific interaction generally is stable under physiological conditions, including, for example, conditions that occur in a living individual such as a human or other vertebrate or invertebrate, as well as conditions that occur in a cell culture such as used for maintaining

mammalian cells or cells from another vertebrate organism or an invertebrate organism. Methods for determining whether two molecules interact specifically are well known and include, for example, equilibrium dialysis, surface plasmon resonance, and the like.

[0281] An agent that alters a specific interaction of a MICAL with a plexin can act, for example, by binding to a MICAL such that it cannot interact specifically with the plexin, by competing with MICAL for binding to the plexin, or by otherwise by-passing the requirement that MICAL specifically interact with a plexin in order to regulate axonal guidance. A mutant plexin that retains its ability to bind to a MICAL but not other plexin functions is an example of an agent that can bind a MICAL, thereby sequestering it and reducing or inhibiting its ability to interact specifically with a functional plexin. A MICAL mutant that includes only its plexin interacting region is an example of an agent that can compete with MICAL for plexin binding, thereby reducing or inhibiting the ability of the MICAL to interact specifically with a plexin. Such MICAL antagonists are useful in practicing a method of the invention, particularly for reducing or inhibiting MICAL axon guidance regulatory activity.

[0282] An agent useful in a method of the invention an antibody that specifically binds a MICAL, including all or a portion of the plexin interacting region, thereby preventing MICAL from interacting specifically with a plexin. Alternatively, the agent can be an antibody that specifically binds to a plexin, including all or a portion of the MICAL interacting region, thereby preventing the plexin from interacting specifically with a MICAL. Such an anti-MICAL or anti-plexin antibody can be selected for its ability to specifically bind MICAL or plexin, respectively, without activating MICAL axon guidance regulatory activity, and can be useful as a MICAL antagonist for reducing or inhibiting MICAL axon guidance regulatory activity; or can be selected for its ability to specifically bind MICAL or plexin, respectively and activate axon guidance regulatory activity, thus acting as a MICAL agonist. The antibody can be raised using a MICAL or a plexin, including plexin A as an immunogen, or can be an anti-idiotype antibody, which is raised against an anti-MICAL antibody and mimics MICAL.

[0283] An agent useful in a method of the invention also can be an agent that reduces MICAL monooxygenase activity, thereby reducing or inhibiting MICAL axon guidance regulatory activity.

[0284] In addition, an agent useful in a method of the invention can be a mutant plexin, which, for example, lacks semaphorin signal transduction activity in response to MICAL binding, or has constitutive semaphorin signal transduction activity. For example, a mutant plexin can have a point mutation, a deletion, or the like in a functional domain other than the MICAL binding domain. Such a dominant negative mutant plexin lacks the ability to transmit a semaphorin and/or MICAL signal despite the fact that it can specifically bind a MICAL.

[0285] An agent useful in a method of the invention also can modulate the level or activity of a MICAL.

[0286] The specific interaction of MICAL with plexin A indicates that MICAL axonal guidance regulatory activity can involve components of the semaphorin-plexin mediated repulsive axon guidance pathway. Thus, the Semaphorin repulsive axon guidance pathway provides a target for modulating the effect of MICAL on a cell, and agents that affect the Semaphorin pathway can be useful for modulating MICAL axon guidance regulatory activity.

[0287] Antagonist agents that can reduce or inhibit MICAL axon guidance regulatory activity are exemplified by dominant negative MICAL polypeptides in which the a functional domain other than the plexin-interacting region has been mutated. The mutants include polypeptides that include a plexin-interacting region and no other

[0288] Where the agent that acts intracellularly is a peptide or a polypeptide, it can be contacted with the cell directly, or a polynucleotide encoding the peptide (or polypeptide) can be introduced into the cell and the peptide can be expressed in the cell. It is recognized that some of the peptides useful in a method of the invention are relatively large and, therefore, may not readily traverse a cell membrane. However, various methods are known for introducing a peptide into a cell. The selection of a method for introducing such a peptide into

a cell will depend, in part, on the characteristics of the target cell, into which the polypeptide is to be provided. For example, where the target cells, or a few cell types including the target cells, express a receptor, which, upon binding a particular ligand, is internalized into the cell, the peptide agent can be operatively associated with the ligand. Upon binding to the receptor, the peptide is translocated into the cell by receptor-mediated endocytosis. The peptide agent also can be encapsulated in a liposome or formulated in a lipid complex, which can facilitate entry of the peptide into the cell, and can be further modified to express a receptor (or ligand), as above. The peptide agent also can be introduced into a cell by engineering the peptide to contain a protein transduction domain such as the human immunodeficiency virus TAT protein transduction domain, which facilitates translocation of the peptide into the cell (see Schwarze et al., *Science* 285:1569-1572 (1999), which is incorporated herein by reference; see, also, Derossi et al., *J. Biol. Chem.* 271:18188 (1996)). The target cell also can be contacted with a polynucleotide encoding the peptide or polypeptide agent, which can be expressed in the cell.

[0289] An agent useful in a method of the invention can be a polynucleotide, which can be contacted with or introduced into a cell as described above. Generally, but not necessarily, the polynucleotide is introduced into the cell, where it effects its function either directly, or following transcription or translation or both. For example, as discussed above, the polynucleotide can encode a polypeptide agent, which is expressed in the cell and modulates MICAL activity. Such an expressed polypeptide can be, for example, a mutant MICAL polypeptide, which does not have monooxygenase activity; or can be a mutant plexin. Methods for introducing a polynucleotide into a cell are exemplified below or otherwise known in the art.

[0290] A polynucleotide agent useful in a method of the invention also can be, or can encode, an antisense molecule, a ribozyme or a triplexing agent. For example, the polynucleotide can be (or can encode) an antisense nucleotide sequence such as an antisense MICAL, plexin, or semaphorin sequence, which can act as an antagonist to reduce or inhibit MICAL axon guidance regulatory activity, thereby inhibiting semaphorin-mediated repulsive axon guidance. Such polynucleotides can be contacted directly with a target cell and, upon uptake by the cell, can effect their antisense, ribozyme or triplexing activity; or

can be encoded by a polynucleotide that is introduced into a cell, whereupon the polynucleotide is expressed to produce, for example, an antisense RNA molecule or ribozyme, which effects its activity.

[0291] An antisense polynucleotide, ribozyme or triplexing agent is complementary to a target sequence, which can be a DNA or RNA sequence, for example, messenger RNA, and can be a coding sequence, a nucleotide sequence comprising an intron-exon junction, a regulatory sequence such as a Shine-Delgarno sequence, or the like. The degree of complementarity is such that the polynucleotide, for example, an antisense polynucleotide, can interact specifically with the target sequence in a cell. Depending on the total length of the antisense or other polynucleotide, one or a few mismatches with respect to the target sequence can be tolerated without losing the specificity of the polynucleotide for its target sequence. Thus, few if any mismatches would be tolerated in an antisense molecule consisting, for example, of 20 nucleotides, whereas several mismatches will not affect the hybridization efficiency of an antisense molecule that is complementary, for example, to the full length of a target mRNA encoding a cellular polypeptide. The number of mismatches that can be tolerated can be estimated, for example, using well known formulas for determining hybridization kinetics (see Sambrook et al., *supra*, 1989) or can be determined empirically using methods as disclosed herein or otherwise known in the art, particularly by determining that the presence of the antisense polynucleotide, ribozyme, or triplexing agent in a cell decreases the level of the target sequence or the expression of a polypeptide encoded by the target sequence in the cell.

[0292] A polynucleotide useful as an antisense molecule, a ribozyme or a triplexing agent can inhibit translation or cleave the nucleic acid molecule, thereby modulating MYCAL axon guidance regulatory activity in a cell. An antisense molecule, for example, can bind to an mRNA to form a double stranded molecule that cannot be translated in a cell. Antisense oligonucleotides of at least about 15 to 25 nucleotides are preferred since they are easily synthesized and can hybridize specifically with a target sequence, although longer antisense molecules can be expressed from a polynucleotide introduced into the target cell. Specific nucleotide sequences useful as antisense molecules can be identified using well known methods, for example, gene walking methods (see, for example, Seimiya et al.,

J. Biol. Chem. 272:4631-4636 (1997), which is incorporated herein by reference). Where the antisense molecule is contacted directly with a target cell, it can be operatively associated with a chemically reactive group such as iron-linked EDTA, which cleaves a target RNA at the site of hybridization. A triplexing agent, in comparison, can stall transcription (Maher et al., *Antisense Res. Devel.* 1:227 (1991); Helene, *Anticancer Drug Design* 6:569 (1991)). Thus, a triplexing agent can be designed to recognize, for example, a sequence of a MICAL gene regulatory element, thereby reducing or inhibiting the expression of a MICAL polypeptide in the cell, and modulating MICAL axon guidance regulatory activity in a target cell.

[0293] The agent to be administered to the subject is administered under conditions that facilitate contact of the agent with the target cell and, if appropriate, entry into the cell. Entry of a polynucleotide agent into a cell, for example, can be facilitated by incorporating the polynucleotide into a viral vector that can infect the cells. If a viral vector specific for the cell type is not available, the vector can be modified to express a receptor (or ligand) specific for a ligand (or receptor) expressed on the target cell, or can be encapsulated within a liposome, which also can be modified to include such a ligand (or receptor). A polypeptide agent can be introduced into a cell by various methods, including, for example, by engineering the peptide to contain a protein transduction domain such as the human immunodeficiency virus TAT protein transduction domain, which can facilitate translocation of the peptide into the cell (see Schwarze et al., *supra*, 1999; Derossi et al., *supra*, 1996).

[0294] The presence of the agent in the target cell can be identified directly, for example, by operatively linking a detectable label to the agent, by using an antibody specific for the agent, particularly a polypeptide agent, or by detecting a downstream effect due to the agent, for example, decreased semaphorin-mediated axon repulsion in the cell. An agent can be labeled so as to be detectable using methods well known in the art (Hermanson, "Bioconjugate Techniques" (Academic Press 1996), which is incorporated herein by reference; see, also, Harlow and Lane, *supra*, 1988). For example, a peptide or polynucleotide agent can be labeled with various detectable moieties including a radiolabel, an enzyme such as alkaline phosphatase, biotin, a fluorochrome, and the like. Where the

agent is contained in a kit, the reagents for labeling the agent also can be included in the kit, or the reagents can be purchased separately from a commercial source.

[0295] An agent useful in a method of the invention can be administered to the site of the pathologic condition, or can be administered by any method that provides the target cells with the polynucleotide or peptide. As used herein, the term "target cells" typical means an immune cell, a transformed eukaryotic cell, or a cell of the nervous system, for example a neuron, that are to be contacted with the agent. For administration to a living subject, the agent generally is formulated in a pharmaceutical composition suitable for administration to the subject. Thus, the invention provides pharmaceutical compositions containing an agent, which is useful for modulating MICAL axonal guidance regulatory activity in a cell, in a pharmaceutically acceptable carrier. As such, the agents are useful as medicaments for treating a subject suffering from a pathological condition as defined herein.

[0296] Pharmaceutically acceptable carriers are well known in the art and include, for example, aqueous solutions such as water or physiologically buffered saline or other solvents or vehicles such as glycols, glycerol, oils such as olive oil or injectable organic esters. A pharmaceutically acceptable carrier can contain physiologically acceptable compounds that act, for example, to stabilize or to increase the absorption of the conjugate. Such physiologically acceptable compounds include, for example, carbohydrates, such as glucose, sucrose or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low molecular weight proteins or other stabilizers or excipients. One skilled in the art would know that the choice of a pharmaceutically acceptable carrier, including a physiologically acceptable compound, depends, for example, on the physico-chemical characteristics of the therapeutic agent and on the route of administration of the composition, which can be, for example, orally or parenterally such as intravenously, and by injection, intubation, or other such method known in the art. The pharmaceutical composition also can contain a second reagent such as a diagnostic reagent, nutritional substance, toxin, or therapeutic agent, for example, a cancer chemotherapeutic agent.

[0297] The agent can be incorporated within an encapsulating material such as into an oil-in-water emulsion, a microemulsion, micelle, mixed micelle, liposome, microsphere or other polymer matrix (see, for example, Gregoriadis, *Liposome Technology*, Vol. 1 (CRC

Press, Boca Raton, FL 1984); Fraley, et al., *Trends Biochem. Sci.*, 6:77 (1981), each of which is incorporated herein by reference). Liposomes, for example, which consist of phospholipids or other lipids, are nontoxic, physiologically acceptable and metabolizable carriers that are relatively simple to make and administer. "Stealth" liposomes (see, for example, U.S. Patent Nos. 5,882,679; 5,395,619; and 5,225,212, each of which is incorporated herein by reference) are an example of such encapsulating materials particularly useful for preparing a pharmaceutical composition useful for practicing a method of the invention, and other "masked" liposomes similarly can be used, such liposomes extending the time that the therapeutic agent remain in the circulation. Cationic liposomes, for example, also can be modified with specific receptors or ligands (Morishita et al., *J. Clin. Invest.*, 91:2580-2585 (1993), which is incorporated herein by reference). In addition, a polynucleotide agent can be introduced into a cell using, for example, adenovirus-polylysine DNA complexes (see, for example, Michael et al., *J. Biol. Chem.* 268:6866-6869 (1993), which is incorporated herein by reference).

[0298] The route of administration of a pharmaceutical composition containing an agent that alters MICAL axon guidance regulatory activity such as semaphorin-mediated axon repulsion activity, will depend, in part, on the chemical structure of the molecule. Polypeptides and polynucleotides, for example, are not particularly useful when administered orally because they can be degraded in the digestive tract. However, methods for chemically modifying polypeptides, for example, to render them less susceptible to degradation by endogenous proteases or more absorbable through the alimentary tract are well known (see, for example, Blondelle et al., *supra*, 1995; Ecker and Crook, *supra*, 1995). In addition, a peptide agent can be prepared using D-amino acids, or can contain one or more domains based on peptidomimetics, which are organic molecules that mimic the structure of peptide domain; or based on a peptoid such as a vinylogous peptoid.

[0299] A pharmaceutical composition as disclosed herein can be administered to an individual by various routes including, for example, orally or parenterally, such as intravenously, intramuscularly, subcutaneously, intraorbitally, intracapsularly, intraperitoneally, intrarectally, intracisternally or by passive or facilitated absorption through the skin using, for example, a skin patch or transdermal iontophoresis, respectively.

Furthermore, the pharmaceutical composition can be administered by injection, intubation, orally or topically, the latter of which can be passive, for example, by direct application of an ointment, or active, for example, using a nasal spray or inhalant, in which case one component of the composition is an appropriate propellant. Furthermore, the agent can be delivered by intrathecal administration using a pump to administer the agent over a period of time.

[0300] A pharmaceutical composition also can be administered to the site of a pathologic condition, for example, intravenously or intra-arterially into a blood vessel supplying a tumor, or by direct injection into the central nervous system., or a portion thereof such as the spinal cord. In aspects of the invention wherein the agent is intended to be delivered to the spinal cord, the agent can be an agent that is capable of crossing into the spinal cord from the blood stream. Such agents include anti-oxidant flavonoids discussed herein.

[0301] The total amount of an agent to be administered in practicing a method of the invention can be administered to a subject as a single dose, either as a bolus or by infusion over a relatively short period of time, or can be administered using a fractionated treatment protocol, in which multiple doses are administered over a prolonged period of time. One skilled in the art would know that the amount of the pharmaceutical composition to treat a pathologic condition in a subject depends on many factors including the age and general health of the subject as well as the route of administration and the number of treatments to be administered. In view of these factors, the skilled artisan would adjust the particular dose as necessary. In general, the formulation of the pharmaceutical composition and the routes and frequency of administration are determined, initially, using Phase I and Phase II clinical trials.

[0302] In general, in methods of the present invention, an agent is administered in an amount that is sufficient to modulate axonal guidance regulatory activity, monooxygenase activity, or plexin-interacting activity. It will be recognized that routine methods can be used to identify effective amounts.

[0303] The pharmaceutical composition can be formulated for oral formulation, such as a tablet, or a solution or suspension form; or can comprise an admixture with an organic or

inorganic carrier or excipient suitable for enteral or parenteral applications, and can be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, or other form suitable for use. The carriers, in addition to those disclosed above, can include glucose, lactose, mannose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, medium chain length triglycerides, dextrans, and other carriers suitable for use in manufacturing preparations, in solid, semisolid, or liquid form. In addition auxiliary, stabilizing, thickening or coloring agents and perfumes can be used, for example a stabilizing dry agent such as triulose (see, for example, U.S. Patent No. 5,314,695).

[0304] In certain embodiments, the present invention provides detection methods. For example, in one embodiment, the present invention provides a method of detecting an immune disease or disorder in a subject. The method includes determining in the subject, the level of expression of a polynucleotide and/or polypeptide of the present invention, such as a MICAL polynucleotide or a MICAL polypeptide. An increased or a decreased level of expression or activity can be indicative of the immune disease or disorder.

[0305] In another embodiment, the present invention provides a method of detecting cancer in a subject. The method includes determining in the subject, the level of expression of a polynucleotide and/or polypeptide of the present invention, such as a MICAL polynucleotide or a MICAL polypeptide. An increased level of expression or activity is indicative of the cancer.

[0306] In another method, the present invention provides a method of assessing the invasiveness of cancer cells in a subject. The method includes determining in the subject, or, more particularly, in cancer cells from the subject, the level of expression of a polynucleotide and/or polypeptide of the present invention, such as a MICAL polynucleotide or a MICAL polypeptide. An increased level of expression or activity indicates that the cancer cells are invasive.

[0307] In another embodiment, the present invention provides a method of detecting central nervous system injury and/or peripheral nervous system injury in a subject, that includes determining the level of expression of a polynucleotide of the present invention, or a polypeptide of the present invention or a MICAL activity thereof, in a sample from the central nervous system or peripheral nervous system of the subject. It has been found that MICAL expression is increased upon injury to the nervous system. Accordingly, an increased level of expression or activity identified by a method of the invention is indicative of the central nervous system injury and/or peripheral nervous system.

[0308] A sample of the central nervous system can be obtained using known methods. The sample can include for example, spinal fluid or neurons from the spinal cord at a suspected site of injury. The subject can be a human. The increased level can be identified by comparing the determined level to a level of a subject not suspected of suffering from spinal cord injury.

[0309] The MICAL activity can be any of the MICAL activities, including, for example, monooxygenase activity, axon guidance regulatory activity, plexin interacting activity, or binding to SH-3 domain-containing proteins. Methods for detecting these activity, some of which are provided herein, are known in the art. Furthermore, methods for determining the level of expression of a polynucleotide or a polypeptide of the present invention, examples of which are provided below, are known in the art.

[0310] Cells obtained in the sample for any of the methods for detecting or assessing of the present invention can be contacted with a lysis buffer. The sample obtained can then be further processed, for example to isolate nucleic acids or polypeptides.

[0311] Nucleic can be isolated from the lysed cells and cellular material by any number of means well known in the art. For example, a number of commercial products are available for isolating polynucleotides, including but not limited to, TriReagent (Molecular Research Center, Inc, Cincinnati, OH). The isolated polynucleotides can then be assayed for the presence of a polynucleotide that encodes a MICAL or MICAL-Like polypeptide.

[0312] Analyzing expression of a MICAL polypeptide or a nucleotide encoding a MICAL polypeptide includes any qualitative or quantitative method for detecting expression of a gene, many of which are known in the art. Non-limiting methods for analyzing polynucleotides and polypeptides are discussed below.

[0313] The methods of analyzing expression of MICAL or a MICAL-Like polypeptide of the present invention can utilize a biochip, or other miniature high-throughput technology. The manufacture and use of biochips such as those involving bioarrays, are known in the art and commercially available (See e.g., bioarrays available from Sigma-Genosys (The Woodlands, TX); Affymetrix (Santa Clara, CA), and Full Moon Biosystems (Sunnyvale, CA)) (For reviews of Biochips and bioarrays *see, e.g.,* Kallioniemi O.P., "Biochip technologies in cancer research," *Ann Med,* Mar; 33(2):142-7 (2001); and Rudert F., "Genomics and proteomics tools for the clinic," Curr Opin. Mol. Ther., Dec;2(6):633-42 (2000)).

[0314] Such bioarrays can be analyzed using blotting techniques similar to those discussed below for conventional techniques of detecting polynucleotides and polypeptides. Other microfluidic devices and methods for analyzing gene expression can be used for the methods of the present invention.

[0315] Quantitative measurement of expression levels using bioarrays is also known in the art, and typically involve a modified version of a traditional method for measuring expression as described herein. For example, such quantitation can be performed by measuring a phosphor image of a radioactive-labeled probe binding to a spot of a microarray, using a phospohor imager and imaging software.

[0316] A method of the present invention in certain aspects, employs RNA, including messenger RNA (mRNA), isolated from a CNS sample. The RNA may be single stranded or double stranded. Enzymes and conditions optimal for reverse transcribing the template to DNA well known in the art can be used. Alternatively, the RNA can be subjected to RNAse protection assays. A DNA-RNA hybrid that contains one strand of each can also be used. A mixture of polynucleotides can also be employed, or the polynucleotides produced in a previous amplification reaction, using the same or different primers may be so used. In

the instance where the polynucleotide sequence is to be amplified the polynucleotide sequence may be a portion of a MICAL, or can be present initially as a discrete molecule, such that the specific sequence is the entire nucleic acid. It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture.

[0317] In addition, RNAse protection assays can be used if RNA is the polynucleotide obtained from the sample. In this procedure, a labeled antisense RNA probe is hybridized to the complementary polynucleotide in the sample. The remaining unhybridized single-stranded probe is degraded by ribonuclease treatment. The hybridized, double stranded probe is protected from RNAse digestion. After an appropriate time, the products of the digestion reaction are collected and analyzed on a gel (see for example Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, section 4.7.1 (1987)). As used herein, "RNA probe" refers to a ribonucleotide capable of hybridizing to RNA in a sample of interest. Those skilled in the art will be able to identify and modify the RNAse protection assay specific to the polynucleotide to be measured, for example, probe specificity may be altered, hybridization temperatures, quantity of nucleic acid etc. Additionally, a number of commercial kits are available, for example, RiboQuantTM Multi-Probe RNAse Protection Assay System (Pharmingen, Inc., San Diego, CA).

[0318] In another embodiment, the polynucleotide in the sample may be analyzed by a blotting procedure, typically a Northern blot procedure, as illustrated in the Examples herein. For blotting procedures polynucleotides are separated on a gel and then probed with a complementary polynucleotide to the sequence of interest. For example, RNA is separated on a gel transferred to nitrocellulose and probed with complementary DNA that is derived from a MICAL gene. The complementary probe may be labeled radioactively, chemically etc. Hybridization of the probe is indicative of the expression of the MICAL.

[0319] Detection of a polynucleotide encoding a MICAL can be performed by standard methods such as size fractionating the nucleic acid. Methods of size fractionating the DNA and RNA are well known to those of skill in the art, such as by gel electrophoresis, including polyacrylamide gel electrophoresis (PAGE). For example, the gel may be a denaturing 7 M or 8 M urea-polyacrylamide-formamide gel. Size fractionating the nucleic

acid may also be accomplished by chromatographic methods known to those of skill in the art.

[0320] The detection of polynucleotides may optionally be performed by using radioactively labeled probes. Any radioactive label may be employed which provides an adequate signal. Other labels include ligands, colored dyes, and fluorescent molecules, which can serve as a specific binding pair member for a labeled ligand, and the like. The labeled preparations are used to probe for a polynucleotide by the Southern or Northern hybridization techniques, for example. Nucleotides obtained from samples are transferred to filters that bind polynucleotides. After exposure to the labeled polynucleotide probe, which will hybridize to nucleotide fragments containing target nucleic acid sequences, the binding of the radioactive probe to target nucleic acid fragments is identified by autoradiography (see *Genetic Engineering,* 1 ed. Robert Williamson, Academic Press (1981), pp. 72-81). The particular hybridization technique is not essential to the invention. Hybridization techniques are well known or easily ascertained by one of ordinary skill in the art. As improvements are made in hybridization techniques, they can readily be applied in the method of the invention.

[0321] Probes according to the present invention and used in a method of the present invention selectively hybridize to a polynucleotide encoding a MICAL polypeptide. In preferred aspects, the probes are spotted on a bioarray using methods known in the art.

[0322] The polynucleotides encoding a MICAL may be amplified before they are detected. The term "amplified" refers to the process of making multiple copies of the nucleic acid from a single polynucleotide molecule. The amplification of polynucleotides can be carried out *in vitro* by biochemical processes known to those of skill in the art. The amplification agent may be any compound or system that will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA polymerase I, Taq polymerase, Klenow fragment of *E. coli* DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, polymerase muteins, reverse transcriptase, ligase, and other enzymes, including heat-stable enzymes (*i.e.*, those enzymes that perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation). Suitable

enzymes will facilitate combination of the nucleotides in the proper manner to form the primer extension products that are complementary to each mutant nucleotide strand. Generally, the synthesis will be initiated at the 3'-end of each primer and proceed in the 5'-direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be amplification agents, however, that initiate synthesis at the 5'-end and proceed in the other direction, using the same process as described above. In any event, the method of the invention is not to be limited to the embodiments of amplification described herein.

[0323] One method of *in vitro* amplification, which can be used according to this invention, is the polymerase chain reaction (PCR) described in U.S. Patent Nos. 4,683,202 and 4,683,195. The term "polymerase chain reaction" refers to a method for amplifying a DNA base sequence using a heat-stable DNA polymerase and two oligonucleotide primers, one complementary to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (-)-strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. The polymerase chain reaction is used to detect the presence of polynucleotides encoding cytokines in the sample. Many polymerase chain methods are known to those of skill in the art and may be used in the method of the invention. For example, DNA can be subjected to 30 to 35 cycles of amplification in a thermocycler as follows: 95°C for 30 sec, 52° to 60°C for 1 min, and 72°C for 1 min, with a final extension step of 72°C for 5 min. For another example, DNA can be subjected to 35 polymerase chain reaction cycles in a thermocycler at a denaturing temperature of 95°C for 30 sec, followed by varying annealing temperatures ranging from 54-58°C for 1 min, an extension step at 70°C for 1 min and a final extension step at 70°C.

[0324] The primers for use in amplifying the polynucleotides of the invention may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof so long as the primers are capable of hybridizing to the polynucleotides of interest. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. The

exact length of primer will depend on many factors, including temperature, buffer, and nucleotide composition. The primer must prime the synthesis of extension products in the presence of the inducing agent for amplification.

[0325] Primers used according to the method of the invention are complementary to each strand of nucleotide sequence to be amplified. The term "complementary" means that the primers must hybridize with their respective strands under conditions, which allow the agent for polymerization to function. In other words, the primers that are complementary to the flanking sequences hybridize with the flanking sequences and permit amplification of the nucleotide sequence. Preferably, the 3' terminus of the primer that is extended has perfectly base paired complementarity with the complementary flanking strand. Primers and probes for polynucleotides encoding MICALs of the present invention, can be developed using known methods combined with the present disclosure.

[0326] Those of ordinary skill in the art will know of various amplification methodologies that can also be utilized to increase the copy number of target nucleic acid. The polynucleotides detected in the method of the invention can be further evaluated, detected, cloned, sequenced, and the like, either in solution or after binding to a solid support, by any method usually applied to the detection of a specific nucleic acid sequence such as another polymerase chain reaction, oligomer restriction (Saiki *et al.*, *Bio/Technology 3*:1008-1012 (1985)), allele-specific oligonucleotide (ASO) probe analysis (Conner *et al., Proc. Natl. Acad. Sci. USA 80*: 278 (1983), oligonucleotide ligation assays (OLAs) (Landegren *et al., Science 241*:1077 (1988)), RNAse Protection Assay and the like. Molecular techniques for DNA analysis have been reviewed (Landegren *et al, Science 242*: 229-237 (1988)). Following DNA amplification, the reaction product may be detected by Southern blot analysis, without using radioactive probes. In such a process, for example, a small sample of DNA containing the polynucleotides obtained from the tissue or subject are amplified, and analyzed via a Southern blotting technique. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. In one embodiment of the invention, one nucleoside triphosphate is radioactively labeled, thereby allowing direct visualization of the amplification product by autoradiography. In another embodiment, amplification primers are fluorescently labeled and run through an

electrophoresis system. Visualization of amplified products is by laser detection followed by computer assisted graphic display, without a radioactive signal.

[0327] The methods of the present invention can involve a real-time quantitative PCR assay, such as a Taqman® assay (Holland et al., *Proc. Natl. Acad. Sci. U S A,* 88(16):7276 (1991)). The assays can be performed on an instrument designed to perform such assays, for example those available from Applied Biosystems (Foster City, CA). Primers and probes for such an assay can be designed according to known procedures in the art.

[0328] Simple visualization of a gel containing the separated products may be utilized to analyze polynucleotides encoding MICALs according to the methods of the present invention. For example, staining of a gel to visualize separated polynucleotides, a number of stains are well known to those skilled in the art. However, other methods known to those skilled in the art may also be used, for example scanning densitometry, computer aided scanning and quantitation as well as others.

[0329] The method for detecting MICAL expression can alternatively employ the detection of a polypeptide product of one of these genes. The method for detecting a MICAL polypeptide in a cell is useful for detecting spinal cord injury by measuring the level of the MICAL polypeptide, in cells obtained from a subject suspected of having, or at risk of having spinal cord injury. The levels of MICALs are indicative of spinal cord injury when compared to a MICAL levels in a subject without spinal cord injury

[0330] In this regard, the sample, as described herein, can be used as a source to isolate polypeptides. The MICAL polypeptide can then be quantified using methods known to those of skill in the art, for example by ELISA.

[0331] Monoclonal antibodies to a particular polypeptide can be used in immunoassays, such as in liquid phase or bound to a solid phase carrier, to detect the MICAL polypeptide In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. Examples of types of immunoassays that can utilize monoclonal antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the

sandwich (immunometric) assay. Detection of the polypeptide antigens using the monoclonal antibodies of the invention can be done utilizing immunoassays, which are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation. In addition, there are a number of commercially available antibodies to cytokines of interest.

[0332] The term "immunometric assay" or "sandwich immunoassay" includes simultaneous sandwich, forward sandwich and reverse sandwich immunoassays. These terms are well understood by those skilled in the art. Those of skill will also appreciate that antibodies according to the present invention will be useful in other variations and forms of assays which are presently known or which may be developed in the future. These are intended to be included within the scope of the present invention.

[0333] Monoclonal antibodies can be bound to many different carriers and used to detect the presence of a MICAL polypeptide. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such using routine experimentation.

[0334] In performing the assays it may be desirable to include certain "blockers" in the incubation medium (usually added with the labeled soluble antibody). The "blockers" are added to assure that non-specific proteins, proteases, or anti-heterophilic immunoglobulins to anti-cytokine immunoglobulins present in the experimental sample do not cross-link or destroy the antibodies on the solid phase support, or the radiolabeled indicator antibody, to yield false positive or false negative results. The selection of "blockers" therefore may add substantially to the specificity of the assays.

[0335] Alternatively, the level of MICAL protein or nucleic acid can be determined *in vivo* using known imaging techniques. For example, an anti-MICAL antibody can be

labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[0336] The results present herein reveal a series of embodiments directed at methods of treating neurological disorders involving a failure of axon regrowth and methods for inducing regrowth of injured processes of neurons that include altering the oxidative state of an affected cell. The present disclosure identifies a novel gene family, the MICALs (also referred to as the zeyphyrins, and the 151 family), whose protein domains suggest a novel means by which nerve growth is regulated. MICALs are characterized by a flavoprotein monooxygenases region (i.e., oxidoreducatase family) that has not been previously shown to function in axon guidance. The presence and necessity of this flavoprotein monooxygenase domain indicates that this gene family regulates repulsive axon guidance through a novel means-oxidation/reduction (redox) mechanisms. The present disclosure indicates that MICALs use redox mechanisms to regulate axon growth by directly or indirectly destabilizing the cellular machinery (i.e., the actin cytoskeleton) necessary for axon outgrowth.

[0337] The discovery of MICALs and their mechanism of action (i.e., redox mechanisms) illuminates a novel general means through which widespread inhibition of axon growth can occur-actin oxidation. In particular, the inability of axons to regrow after spinal cord injury may be a consequence of the presence of high amounts of reactive oxygen species and other oxidative mechanisms in the spinal cord milieu after injury. These oxidants may directly alter the structure of the actin cytoskeleton-in effect, acting non-specifically like members of the MICAL protein family. In total, these discoveries indicate novel and immediate treatments for many neurological disorders, targeting both general oxidants as well as the MICALs. In particular, these novel treatments include novel therapeutic strategies (e.g., antioxidants and other redox active compounds) and novel agents (e.g., EGCG, EC, and other flavonoids and antioxidants) to promote axonal regrowth (e.g., following spinal cord injury and similar neurological disorders) as well as novel strategies (e.g., oxidants) to limit abnormal and excessive axonal growth (e.g., following certain neuropathies, and increased sensitizations). In summary, the discoveries disclosed herein, implicate oxidation mechanisms in limiting axon growth and preventing axon

regeneration in general; these mechanisms have not previously been suggested or shown to be involved in limiting axon outgrowth after spinal cord injury or after other neurological disorders.

[0338] Accordingly, in another aspect, the present invention provides a method for treating a neurological disorder involving a failure of axon regrowth, comprising contacting a neuron having axons that fail to regrow, or surrounding tissue, with an agent that neutralizes oxidants, thereby treating the neurological disorder. The surrounding tissue can include any tissue whose components, typically cells, can produce factors that affect axonal growth.

[0339] The agent that neutralizes oxidants can include virtually any anti-oxidant. Especially preferred are anti-oxidants that can be delivered orally and that can enter the central nervous system or peripheral nervous system. The agent for example, can be applied directly to a neuron having axons that fail to regrow. For example, the agent can be directly injected to the site of spinal cord injury. A method according to this aspect of the invention can be performed *in vitro* or *in vivo*.

[0340] For example, anti-oxidant vitamins can be used. These vitamins include vitamins E, C and beta carotene. Other useful antioxidants for the present invention include, for example, methylprednisolone, Tirilazad, lazaroids (21-aminosteroids) and similar steroids, alpha tocophenol, lycopene, gamma tocophenol, mannitol, catalase, and glutathione, superoxide dismutase. Also useful for the present invention is the anti-oxidant compound H 290/5 (See e.g., Thornwall M., et al., *Acta Neurochir Suppl (Wien)* 70:212-5 (1997)), and the anti-oxidant AM-36 (Callaway, J. K., *J. Alzheimers Dis.* 2(2):69-78 (2000)).

[0341] As indicated herein, monooxygenase inhibitors and anti-oxidant flavonoids, including those that are monooxygenase inhibitors such as ECGC and EC, can be used as anti-oxidants for this embodiment and embodiments aimed at inducing regrowth of an injured process of a neuron. The anti-oxidant flavonoid can be a gallic acid derivative such as ECGC or EC. Examples of other gallic acid derivatives that can be used to affect axonal guidance regulatory activity in embodiments of the present invention include, but are not limited to, (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-

epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3"-*O*-methyl-EGCG, 3"-*O*-methyl-ECG, 3"-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, and *n*-cetyl gallate.

[0342] In another embodiment, the present invention provides a method for inducing regrowth and/or preventing inhibition of an injured process of a neuron, that includes altering the levels of reactive oxygen species in the milieu of the neuron. The method can include identifying a site that includes the neuron suspected of having an injured process, before altering the levels of reactive oxygen species or other oxidation products in the milieu of the neuron. The neuronal process can be an axon or a dendrite. The levels of reactive oxygen species or other oxidation products are typically decreased by the method.

[0343] In certain aspects, levels of reactive oxygen species are altered chronically, as discussed herein for application of an agent. For example, the levels can be altered for a period of time that is sufficient to permit axon regrowth (i.e. neurorestoration) and the establishment of synaptic connections with new targets. This chronic alteration of the level of reactive oxygen species, in certain aspects is for at least 1, 2, 7, or 14 days, or 1, 2, 3, 4, 5, 6, 12, 24, 36, 48, or 60 months after identification or suspicion of the injured process of the neuron.

[0344] In one aspect, oxygen species can be decreased chronically by delivery of a recombinant cell that expresses a recombinant enzyme that lowers reactive oxygen species to a site of neurological damage or other site in need of regrowth of neural processes. Such recombinant enzymes include, for example, catalase and superoxide dismutase. Alternatively, enzymes that lower reactive oxygen species can be delivered to a site in need of regrowth of neural processes.

[0345] The method can further include adding an agent that promote neuron process regrowth, such as a neurotrophic factor or a neural stem cell, as discussed above, to the milieu of the neuron. The milieu of the neuron includes fluids, molecules, and tissues that surround a neuron. As discussed hereinabove, the agent that promotes neuron process regrowth can be, for example, a neurotrophin, a mechanical bridge, or a stem cell.

[0346] In another embodiment, the present invention provides a method for limiting abnormal axon outgrowth, that includes contacting a neuron or the milieu of the neuron with an agent that affects oxidation state. The abnormal axon outgrowth can be excessive axon outgrowth.

[0347] In another embodiment, the present invention provides a method for improving sperm function, that includes contacting a sperm cell, or progenitor thereof, with an antioxidant agent in an amount sufficient to modulate MICAL activity. In certain aspects, the method includes reducing levels of reactive oxygen species in the milieu of the sperm or progenitor thereof. Human spermatozoa exhibit a capacity to generate ROS and initiate peroxidation of the unsaturated fatty acids in the sperm plasma membrane, which plays a key role in the etiology of male infertility (Sharma R. K., and Agarwal A., *Urology*, 48:835 (1996)). Accordingly, MICAL expression can be involved in sperm malfunction through oxidation of components of sperm cells. Therefore, agents such as antioxidants, for example, antioxidant flavonoids, and particularly monooxygenase inhibitors, such as those disclosed herein can be used to improve sperm function and to treat male infertility.

[0348] In another embodiment, the present invention provides a method for modulating cardiac development in a subject, for example a human subject such as a patient in need of the method. The method includes contacting a cardiac neural crest cell with an amount of an agent that modulates MICAL activity, the amount being effective to modulate cardiac development.

[0349] In another embodiment, the present invention provides a method for treating, managing, and/or ameliorating the symptoms of a cardiovascular disease in a human subject. The method includes contacting a cardiac cell in the human subject with an amount of an agent that modulates MICAL activity. The amount is effective to treat, manage, and/or ameliorate the symptoms of the cardiovascular disease. The agent can modulate any of the activities of a MICAL included, for example, axon guidance regulatory activity, plexin interacting activity, actin binding, and/or monooxygenase activity.

[0350] In another embodiment, the present invention provides a method for modulating an immune response in a human subject in need thereof. The method includes contacting an

immune cell in the human subject with an amount of an agent that modulates MICAL activity. The amount is effective for modulating said immune response. In certain aspects, the immune response is inflammation. MICALs through their involvement in semaphorin-mediated pathways, are predicted to be involved in semaphorin-mediated processes of the immune system. For example, using a differential display technique, upregulation of semaphorin E in rheumatic synovial fibroblasts has been observed. Accordingly, in certain aspects, the human subject for the method for modulating an immune response has rheumatoid arthritis. In other embodiments, the human subject for the method of modulating an immune response has another inflammatory disease, such as, but not limited to, asthma, encephilitis, inflammatory bowel disease, chronic obstructive pulmonary disease (COPD), allergic disorders, septic shock, pulmonary fibrosis, undifferentitated spondyloarthropathy, undifferentiated arthropathy, inflammatory osteolysis, and chronic inflammation resulting from chronic viral or bacteria infections.

[0351] In another embodiment, the present invention provides a method for inhibiting cancer cell proliferation or metastasis in a human subject. The method includes contacting a cancer cell in the human subject with an amount of an agent that modulates MICAL activity. The amount is effective to inhibit MICAL axon guidance regulatory activity, monooxygenase activity, or plexin-interacting activity, thereby being effective to inhibit proliferation or metastasis of the cancer cell. As discussed herein, MICALs through their involvement in semphorin-mediated pathways are predicted to be involved in semaphorin-mediated processes in cancer cells, including metastatic cancer cells.

[0352] In another aspect, the present invention includes kits that are useful for carrying out the methods of the present invention. The components contained in the kit depend on a number of factors, including the type of method being carried out.

[0353] Accordingly, the present invention provides a kit for modulating the activity of a MICAL polypeptide. The kit includes an agent in an effective amount and formulation to be effectively delivered to a subject. The agent can be any of the agents disclosed herein. In certain aspects, the agent is a monooxygenase inhibitor, such as a flavonoid, for example a gallic acid derivative. The gallic acid derivative in certain aspects is ECGC or EC. The kit also includes instructions, either as a pamphlet provided with the kit, or in an on-line site

that provides instructions, for performing the method for modulating activity of the MICAL polypeptide.

[0354] In another aspect of a kit embodiment, the present invention provides one or more containers that include a MICAL or MICAL-Like polypeptide or polynucleotide of the present invention, a vector that includes the MICAL or MICAL-like polypeptide operably linked to a heterologous promoter, and/or a recombinant cell that includes the vector. The kits can include instructions for performing any of the methods provided herein.

[0355] In another aspect, the kit can provide a container that includes a MICAL detection molecule. A MICAL detection molecule is for example, an antibody, an oligonucleotide probe, or any of the other known types of molecules that can be used to detect expression or activity of MICAL, as disclosed herein. The kit in certain aspects includes an oligonucleotide probe, primer, or primer pair, or combination thereof for carrying out a detection method of the present invention, as discussed above. For example, the probe, primer, or primer pair, can be capable of selectively hybridizing to a MICAL polynucleotide. The kit can further include one or more detectable labels.

[0356] The following examples are intended to illustrate but not limit the invention.

EXAMPLE 1

MICAL is a Large, Cytosolic, Multidomain Protein that Interacts with *Drosophila* Plexin A

[0357] This example illustrates that *Drosophila* MICAL is a multi-domain protein that interacts with Plexin A.

Yeast Two-Hybrid Screening

[0358] Yeast protocols were conducted using standard techniques (Golemis et al., 1994). Portions of the intracellular domains of PlexA (amino acids 1702-1945; EST LD13083), PlexB (amino acids 1785-2051; EST CK00213), and the corresponding intracellular regions

of human Plexin A3, and mouse Plexin A4 (gifts of L. Tamagnone, and H. Fujisawa, respectively) were inserted into the yeast bait vector, as described in more detail as follows:

[0359] The terminal "C2" portion of the PlexA cytoplasmic domain (amino acids 1702-1945), which is highly conserved among all plexin family members, was used to search for interacting proteins encoded by a *Drosophila* embryonic (0-24 hrs.) cDNA library. The PCR-amplified PlexA C2 domain (the bait) was inserted into the yeast expression vector pEG202 (bait vector). Following sequencing of both strands, the bait was introduced into the yeast strain EGY48 containing the β-galactosidase expressing plasmid pJK103. Western analysis of transformed yeast using an antibody to LexA (Invitrogen) confirmed appropriate sized expression and an activation assay showed that the bait could not activate transcription on its own . A 0-24 hr. *Drosophila* embryonic cDNA library was cloned into the yeast expression vector pJG4-5 (generated by H. Araj). > 2 X 10^6 clones were screened and interactions assessed with a visual β -galactosidase assay and a test of growth in the absence of leucine. Yeast clones exhibiting varying degrees of interaction were selected and standard protocols were used to recover the library vector and sequence these clones on both strands.

[0360] cDNAs containing the C-terminal of human MICAL-1 (DFkzp434B517) and mouse MICAL-2 (BB481898) were used to amplify portions homologous to the last 200 amino acids of *Drosophila* MICAL and cloned into the library vector.

Molecular Analysis

[0361] Proteins, domains, and alignments were identified using Web-based protein domain searching and alignment tools (PFAM, BLAST, PRINTS, JALVIEW, and ClustalX) and our own molecular analysis. Human MICAL-1 (EST FLJ11937), Human MICAL-2 (ESTs BF815128, KIAA1364, KIAA0819), and Human MICAL-3 (ESTs KIAA0750, and FLJ14966) were identified by BLAST searches on publicly available cDNA and genomic sequence and in some cases overlapping ESTs were assembled virtually. *Drosophila* MICAL-L (EST LD45758) and human MICAL-L1 (EST XM001070) and MICAL-L2 (ESTS FL00139 and FLJ23471) were identified by searching publicly available cDNA and genomic sequence databases.

RESULTS

[0362] To identify mediators of semaphorin-dependent repulsive axonal guidance the terminal highly conserved "C2" portion of the PlexA cytoplasmic domain was used to search for interacting proteins encoded by a *Drosophila* embryonic (0-24 hrs.) yeast two-hybrid cDNA library (Figures 5A–C). Fifty-two interactors encoded by cDNAs derived from four different genes were identified. Over half of these interactors were encoded by a single gene and two overlapping cDNAs encoded by this gene were selected for further study (clones 23 and 151).

[0363] Yeast interactions using the C2 domain of PlexA as the bait (Plexin A C2) were assessed and the strongest interactors (e.g., clones 23, and 151), as determined by a β-galactosidase assay (Beta Gal Activity), were derived from MICAL (Fig. 5B). Further, the bait construct (Plexin A C2) was cloned into the library vector and clone 151 was inserted into the bait vector and an interaction assay demonstrated the vector independence of these interactions . Clones 23 and 151 do not interact with the C2 domain of *Drosophila* PlexB. Though GOF experiments suggest that PlexB functions like PlexA to signal semaphorin-mediated motor axon repulsion (Hu et al., 2001), we have thus far not observed associations between MICAL and PlexB proteins. The C2 domains of human Plexin A3 (HPlexin A3 C2) and mouse Plexin A4 (MPlexin A4 C2) interact strongly with the plexin interacting regions (PIR) of human MICAL-1 (HMICAL1 PIR) and mouse MICAL-2 (MMICAL2 PIR), respectively.

[0364] DNA sequence analysis suggested that the overlapping cDNA clones we identified in our yeast screen did not encode a full-length gene product. These cDNAs have an open reading frame (ORF) at their 5' ends and a stop codon near their 3' ends, indicating we had identified the C terminal 255 amino acids of this novel protein. Northern analysis using standard techniques (Sambrook et al, 1989) on 0-24 hr *Drosophila* embryonic total RNA and a portion of clone 151 as a probe showed that the full-length transcript from this gene is greater than 10 kilobases (Kb) in length (Clone 151, 3' MICAL) (Fig. 5C). Given a lack of publicly available expressed sequence tags (ESTs) extending our cDNA further 5', we used one of our initial cDNA clones (clone 23; Figure 1D) to screen a *Drosophila* embryonic lambda gt11 phage cDNA library (a generous gift from K. Zinn) for full-length

cDNAs using standard techniques (Sambrook et al., 1989). The longest clones were selected and sequenced on both strands. We were unable to identify a single full-length transcript, so we conducted an extended cDNA walk to obtain full-length MICAL transcripts. Isolated cDNAs were assembled to identify full-length MICAL isoforms. Northern analysis using a probe derived from the 5' end of an assembled full-length cDNA detected a large transcript of greater than 10kb (5' MICAL). This transcript is similar in size to the transcript detected with a probe from the 3' portion of this assembled cDNA, providing further evidence that the 3' and 5' portions of the assembled full-length cDNAs are from the same transcript or group of large transcripts.

[0365] The genomic organization of the *MICAL* locus was determined using the Sequencer 2.1 program (Gene Codes Corp.), the identified cDNAs, and publicly available *Drosophila* genomic DNA sequences. This extensive molecular analysis demonstrated that the *Drosophila* gene defined by clones 23 and 151, is *Drosophila* MICAL, and covers >41kb of genomic sequence and has at least 25 exons (Figure 1A; see Figure 5). Based on analysis of isolated cDNAs and western analysis (see Figure 6D), there are at least three MICAL isoforms ("long," "medium," and "short" variants; Figure 1A).

[0366] At the MICAL C-terminus is the plexin interacting region that was identified in the yeast screen. Within the plexin interacting region there is a predicted heptad-repeat, coiled-coil structure (Figure 1B), a motif thought to be involved in protein-protein interactions (Burkhard et al., 2001). Interestingly, this region of MICAL shares amino acid similarity with several other coiled-coil domain-containing proteins including a portion of the alpha domain found in the Ezrin, Radixin, and Moesin (ERM) proteins (~22% identity; Bretscher et al., 2000). The last four amino acids of MICAL (ESII) are a PDZ protein binding motif (Harris and Lim, 2001). N-terminal to the plexin interacting region of MICAL there is a proline rich region. MICAL has two regions of varying length, variable regions (1) and (2), which have no significant similarity to any other proteins and which appear to determine the size of the different MICAL proteins (Figure 1B). MICAL has a single LIM domain (Figure 1B), a protein-protein interaction module found in a variety of proteins involved in signal transduction cascades and in cytoskeletal organization (Bach, 2000), and also a single calponin homology (CH) domain (Figure 1B), a domain also found

in cytoskeletal and signal transduction proteins and known to be involved in actin filament binding (Gimona et al., 2002). The MICAL N-terminal ~500 amino acid domain is highly conserved among MICAL-related proteins (see below), but is unique over its entire length in comparison to other proteins.

EXAMPLE 2

MICAL IS EXPRESSED ON *DROSOPHILA* EMBRYONIC MOTOR AND CNS AXONS AND COIMMUNOPRECIPITATES WITH PLEX A

[0367] This example illustrates that MICAL is expressed in axons and that MICAL interacts with PlexA.

In Situ Hybridization

[0368] RNA in situ analysis of whole-mount *Drosophila* embryos and cryosections of E15 and E18 rat spinal cords were as described (Kolodkin et al., 1993; Pasterkamp et al., 1998).

Development of HA-PlexA transgenic flies

[0369] The HA-PlexA construct was created by inserting in the correct orientation an in-frame PCR amplified HA sequence into the EcoRI site that links the artificial signal sequence and the extracellular domain of PlexA in a PlexA pSectag B construct generously provided by C. Goodman. Following sequencing of the insert, the entire HA-PlexA cDNA was inserted into the pUAST vector; one transgenic fly was obtained.

MICAL Antibody Generation, Western Analysis, Immunohistochemistry, and Immunoprecipitation

[0370] Antibodies were generated and characterized as described (Yu et al, 1998). cDNAs corresponding to the last 359 amino acid of MICAL (MICAL-CT antibody) were inserted into the pTrcHisA vector (Invitrogen). MICAL-CT antibodies were used for Western analysis at a 1:2000 dilution, and on *Drosophila* embryos at 1:3000 dilution.

[0371] Embryos generated by crossing UAS-HAPlexA and Elav-GAL4/Cyo adults were collected, and co-immunoprecipitations were performed using standard techniques, and an HA monoclonal antibody (12CA5; Roche). Western analysis was performed using an HA antibody (1:3000, rat mAb Clone 3F10, Roche), our MICAL-CT antibody (1:2000), and an Enabled antibody (1:500; IG6C10; gift from D. Van Vactor).

Immunoprecipitations

[0372] Embryos generated by crossing UAS-HAPlexA and Elav-GAL4/Cyo adults were collected, dechlorinated with 50% bleach, and 100-200 mg of embryos were lysed in either 1) 1mL of RIPA buffer (150mM NACl, 1.0% NP-40, 0.5% deoxycholate, 0.1% SDS, 50mM Tris, pH 8.0, 0.2mM $NaVO_4$, 10mM NaF, and protease inhibitor cocktail (Sigma) and 20μg/mL PMSF), 2) 1mL of 1% Triton buffer (150mM NaCl; 50mM Tris-HCl, pH 8.0), or 3) 1% NP-40 (150mM NaCl, 5mM Tris-HCl, pH 8.0) using a tight 2mL Dounce homogenizer at 4°C. Similar results were observed with each buffer. Extracts were cleared by ultra centrifugation at 100,000xg for 15 min. at 4°C and added to 50μL of a 50% slurry of Gammabind G beads (Amersham) for 30min with rocking at 4°C. Lysates with beads were then centrifuged for 30 minutes and supernatants were immunoprecipitated for 30 minutes with 2μL anti-HA per sample (mouse monoclonal antibody (Clone12CA5, Roche). 100μL of Gammabind G beads were then added to the sample, and the samples were incubated for 90 min. at 4°C with rocking, washed 6 times with lysis buffer and resuspended in 50μL of Laemmli loading buffer. Western analysis was performed using an HA antibody (1:3000, rat mAb Clone 3F10, Roche), our MICAL-CT antibody (1:2000), and an Enabled antibody (1:500; IG6C10; a generous gift from D. Van Vactor).

RESULTS

[0373] *In situ* hybridization analysis using RNA probes corresponding to the N- or C-terminal of MICAL shows that MICAL and PlexA have similar patterns of embryonic mRNA expression. During early *Drosophila* development (stages 7-8), both MICAL and PlexA are expressed in the ventral neurogenic region and in many non-neuronal tissues (including developing mesoderm, cells surrounding the cephalic furrow and amnioproctodeal invagination, and in gut primordia). This non-neuronal expression is also

seen later in embryonic development (stages 11–17), where both MICAL and PlexA are present within the anterior and posterior midgut primordia, the visceral musculature, and weakly in somatic musculature. During axonal pathfinding (stage 13 onward) both MICAL and PlexA are expressed within the developing brain and ventral nerve cord in most, if not all, CNS neurons but MICAL, like Sema1a and PlexA, is not highly expressed in peripheral sensory neurons.

[0374] Western blot analysis using a polyclonal antibody directed against the MICAL C-terminus (MICAL-CT) revealed prominent bands at 530kD, 330kD, 300kD, 200kDa, and 125kDa in lysates from wild-type embryos which are seen at greater intensity in lysates from embryos harboring a chromosomal duplication that includes the MICAL locus (see Figure 6). The three largest protein bands are in agreement with the molecular weights of the three MICAL isoforms predicted from our analysis of MICAL cDNAs (Figures 1A). MICAL immunoreactivity was not observed in embryonic lysates obtained from mutant embryos harboring a deficiency which includes the MICAL locus (see below), showing these products identified by Western analysis are derived from MICAL and that our antibodies are MICAL-specific (see Figure 6D).

[0375] MICAL protein is present in neuronal cell bodies, along axons, and in growth cones. MICAL immunostaining first appears in the nervous system at stage 13 and labels motor and CNS projections. At later embryonic stages, MICAL immunostaining is present on axons that make up all motor axon pathways: the intersegmental nerve (ISN); the intersegmental nerves (ISNb and ISNd); and the segmental nerves a and c (SNa, and SNc) . MICAL immunostaining is also present in segment boundaries at the position of muscle attachment sites and at low levels in the lateral cluster of chordotonal organs.

[0376] To ask whether PlexA and MICAL directly interact in neurons, transgenic flies were generated that contain a transgene encoding epitope-tagged PlexA (HA-PlexA) under the control of an upstream activator sequence (UAS) (Brand and Perrimon, 1993) and crossed with flies that express the GAL4 transcription factor in all neurons (Elav-GAL4). Lysates from embryos containing both HA-PlexA and Elav-GAL4 elements were subjected to immunoprecipitation using HA antibodies and then Western blotting with MICAL-CT antibodies. Robust co-immunoprecipitation of MICAL was observed using HA antibodies

and also reciprocal co-immunoprecipitation of HA-PlexA using MICAL-CT antibodies . The "large" MICAL isoform is the predominant variant observed to be associated with neuronally-expressed HA-PlexA, which may reflect tissue-specific expression of this isoform in neurons. MICAL co-immunoprecipitation appears specific since enabled (Ena), a neuronally expressed cytosolic protein, was not co-immunoprecipitated by HA antibodies in similar experiments and Unc5, a neuronally-expressed transmembrane receptor was not co-immunoprecipitated by MICAL-CT antibodies .

EXAMPLE 3

A MICAL LOSS-OF-FUNCTION MUTANT DEMONSTRATES THAT MICAL IS REQUIRED FOR MOTOR AXON PATHFINDING

[0377] This example illustrates that MICAL is required for motor axon pathfinding.

Drosophila Genetics and Phenotypic Characterization

[0378] *Drosophila* genetics, transformations, and preparation and analyses of *Drosophila* embryos was performed as described (Winberg et al., 1998b; Yu et al., 1998). The cytological location of *MICAL* was determined by hybridizing a radiolabeled cDNA probe corresponding to either the 5' or the 3' regions of the *MICAL* ORF on a *Drosophila* genomic P1 clone filter (Genome Systems) and following the manufacturer's instructions. *MICAL* is located on the third chromosome of *Drosophila* in the 85F3-6 chromosomal location . Unfortunately, this region was devoid of any small, publicly available, deficiencies and candidate *MICAL* mutations.

[0379] To generate a *MICAL* LOF mutant we identified two P transposable elements closely flanking the *MICAL* locus and used a P element transposase-mediated mutagenesis strategy to delete the region between these P elements (Cook et al., 2001; Cooley et al., 1990; Preston et al., 1996). A search of public databases revealed two transposable elements that flanked the *MICAL* locus and were separated by ~165 kbs. Our molecular analysis confirmed that one (*l(3)s2681;* Bloomington Stock Center) was located in a separate gene <3kb from the putative 5' end of *MICAL* and the other (*EP(3)3681*; Berkeley

Drosophila Genome Project) was ~120 kb from the *MICAL* 3' end . A third P element (*l(3)10477*) was situated between these two elements and served as a genetic marker. Our molecular analysis shows that *l(3)10477* (Bloomington Stock Center) is situated in a novel gene ~70kb from the 3' end of *MICAL*, between l(3)s2681 and EP(3)3681. *l(3)10477* adult flies hold their wings out-stretched at a 45° angle from their bodies. Molecular analysis, genetic complementation analysis, and identification of additional *l(3)10477* alleles (including on the TM3 balancer) show that the wing phenotype associated with *l(3)10477* is recessive and due to a mutation of a gene located ~70kb from the 3' end of *MICAL* (Terman and Kolodkin, unpublished). We have called this new gene *stretched out* (*stretch*) and used it as a genetic marker to identify *MICAL* deletions (submitted to Flybase). (1) Using standard genetic techniques, we generated adult flies containing each starting P element in trans (*EP(3)3681/l(3)s2681*). (2) A source of P element transposase was introduced as a mutagen and screened for the appearance of adult progeny containing a *stretch* wing phenotype (Screened through ~25,000 flies).

[0380] 80 adults were identified with a stretch-like wing phenotype and individual fly stocks were made for each. Each stock was then re-scored for the *stretch* phenotype. Five candidate *MICAL* deficiency lines were identified and the extent of the deletion was mapped using standard techniques by asking whether these lines complemented ("+") or failed to complement ("-") (i.e., showed lethality or the wing phenotype) when crossed to flies containing chromosomal aberrations flanking the *MICAL* locus. The fly stocks used in the mapping (complementation analysis) experiments were as follows: *Df(3R)by10* (85D8;85E10-13; Bloomington Stock Center), *Df(3R)by62* (85D11-14;85F6, 041h; 85F6(T); Bloomington Stock Center), *Df(3R)segG16* (segregant *Dp(2;2)G16*[2D]TE35B-3[2P]; breakpoints=*Df(3R)85F6-8;85F12-86A2* + *Dp(2;2)35B2;35D7* was made from *T(2;3)G16*; for simplicity referred to as *Df(3R)segG16*; kindly provided by John Roote and Michael Ashburner), and two small deficiencies generated in our lab (*Df(3R)mr73* and *Df(3R)mrO1*; Terman and Kolodkin, unpublished). One of the five lines which exhibited a stretched out wing phenotype and removes *MICAL* ($Df(3R)swp2^{MICAL}$) was used to characterize the *MICAL* loss-of-function phenotype, and the extent of this deletion is indicated in Fig 6C.

Western blot analysis

[0381] Late stage 16/17 *Drosophila* embryos were genotyped and homogenized in lysis buffer (1% NP-40, 50 mM Tris-HCl pH 8.0, 2 mM EDTA, protease inhibitor cocktail), run on 4% SDS-PAGE gel, and subjected to western blotting. Lysates from 5 wild-type embryos, 5 embryos carrying a duplication of the *MICAL* locus (Dp(3;3)M86D[+]2 (85D1-4;87A5; Bloomington Stock Center), and 5 Df(3R)swp2^{MICAL} embryos were blotted with either MICAL polyclonal antisera (MICAL-CT) or as a loading control an enabled monoclonal antibody (IG6C10). Prominent bands are observed at 530 kD, 330 kD, 300 kD, 200 kDa, and 125 kDa in wild type and at stronger intensity in *MICAL* duplication embryo; none of these bands are observed in *Df(3R)swp2*^{MICAL} embryos.

***Drosophila* Transformation Constructs**

[0382] A *MICAL* rescue construct (*UASMICAL*) was physically assembled from isolated cDNAs and cloned into the pUAST vector for *Drosophila* germline transformation (Yu et al, 1998). Three independent transgenics were obtained. In addition to using the rescue construct to attempt to rescue the LOF and the genetic interaction phenotypes, this construct was used to examine the effects of overexpression of MICAL in all neurons. Expressing one copy of *UASMICAL* in all neurons in a wild-type background resulted in less penetrant phenotypes than expressing 2 copies .

RESULTS

[0383] To determine whether MICAL functions *in vivo* to propagate Sema-1a-mediated motor axon guidance, detailed genetic analyses were performed of MICAL gain- and loss-of-function mutants. A small, tractable, deficiency (called Df(3R)swp2MICAL; see Figure 6) was generated that removes ~170Kb that includes the MICAL locus and ~six other genes. Western blot analysis on lysates from embryos homozygous for Df(3R)swp2MICAL demonstrates a loss of all MICAL protein (Figure 6D), and no MICAL immunostaining is observed in these embryos . These data, in combination with rescue experiments using a MICAL cDNA (see below), define the small deficiency

Df(3R)swp2MICAL as a MICAL null allele. Df(3R)swp2MICAL homozygotes survive into larval stages and have no overt morphological abnormalities (see Figure 6).

[0384] If MICAL functions in Sema-1a/PlexA–dependent repulsive axon guidance, then MICAL LOF mutants should exhibit motor axon guidance defects similar to the distinct and highly penetrant defects seen in Sema1a and PlexA LOF mutants. The development of the stereotypic pattern of neuromuscular connectivity in embryonic *Drosophila* abdominal segments is observed with anti-fasciclin II (mAb 1D4) staining of stage 16/17 embryos (VanVactor et al., 1993). Motor axons initially exit the CNS as part of either the intersegmental nerve (ISN) or the segmental nerve (SN). They are then guided into five major nerve branches (the ISN, ISNb, ISNd, SNa, and SNc), each of which targets different muscle groups such that individual motor axons eventually innervate individual target muscles (Landgraf et al., 1997).

[0385] In wild-type embryos, the ISNb is formed by ISN axons defasciculating and extending dorsally through the ventral musculature to innervate muscles 6 and 7 and muscles 12 and 13. Axons within the ISNb pathway in Sema1a or PlexA mutants often fail to defasciculate and innervate their muscle targets (Table 1; Winberg et al., 1998b; Yu et al., 1998). In the absence of MICAL, axons within the ISNb show similar highly penetrant ISNb phenotypes (Table 1). These phenotypes include the failure of some or all axons to defasciculate from the ISN, stalling of axons within the ISNb following defasciculation from the ISN, ISNb axons bypassing their target muscle groups, and greatly reduced or absent innervation of target muscles.

[0386] Axons within the SNa pathway in MICAL mutants also exhibit highly penetrant defects similar to those observed in both Sema1a mutants and PlexA mutants. In wild-type embryos, SNa axons defasciculate from the SN and extend through the ventral musculature as a single tightly fasciculated bundle. At the dorsal edge of muscle 12, SNa axons defasciculate to give rise to a dorsal (D) and lateral (L) branch. Axons within the dorsal branch extend dorsally between muscles 22 and 23 and then make two characteristic turns, continuing further dorsally between muscles 23 and 24 . In Sema1a and PlexA mutants, SNa axons within the dorsal branch often stall near muscle 12 and fail to reach the dorsal-most portion of their trajectory (Table 1; Winberg et al., 1998b; Yu et al., 1998). MICAL

mutants exhibit similar, highly penetrant, SNa stall phenotypes (Table 1). MICAL mutants also exhibit prominent guidance defects in axons that give rise to the ISNd, SNc, TN, and the third most lateral fasciclin II -positive CNS longitudinal connective-defects, which have been observed in Sema1a and PlexA mutants (Winberg et al, 1998b; Yu et al, 1998). In MICAL LOF mutant embryos additional phenotypes beyond those seen in PlexA and Sema1a mutants were not observed, suggesting that MICAL primarily functions during *Drosophila* neural development in PlexA signaling events.

[0387] MICAL expression was restored in homozygous Df(3R)swp2^{MICAL} embryos using one copy of the transgenic construct UAS- ^{MICAL} under the control of the neuron-specific driver Elav-GAL4. Due to the large size of the MICAL protein, rescue was attempted using the smallest MICAL isoform—the 300kD "small" form (Figure 1A). The level of neuronal MICAL expression observed by immunostaining with MICAL antibodies in three independent MICAL transformants was somewhat lower than that seen in wild type embryos . Neuronal MICAL expression did not rescue the adult lethality in Df(3R)swp2 ^{MICAL} homozygotes, suggesting a requirement for the MICAL "long" form, other genes within the Df(3R)swp2 ^{MICAL} deficiency, and/or MICAL in non-neuronal cells for adult viability. We did, however, observe that neuronal MICAL expression in homozygous Df(3R)swp2 ^{MICAL} embryos almost completely rescues embryonic ISNb and SNa motor axon guidance defects (Table 1) and CNS longitudinal connective defects . Therefore, axon guidance phenotypes observed in the MICAL deficiency Df(3R)swp2 ^{MICAL} result from a lack of neuronal MICAL.

Table 1. Axon guidance phenotypes (ISNb and SNa Phenotypes)

Genotype (hemisegments)	Abnormal ISNb Bypass[a](x)	Abnormal Muscle 6/7 Innervation[b](y)	Abnormal Muscle 12/13 Innervation	SNa Pathway[c](z)
CONTROLS:				
+/+ (wild type) (n=120))	0%	1.7%	2.5%	10.0%
Elav-GAL4/+ (n=130)	0% (0%)	1.5% (0%)	4.6%	8.5% (0%)
Elav-GAL4/Elav-GAL4 (n=109)	0% (0%)	2.8% (0%)	8.2%	12.8% (0%)
Df(3R)swp2^{MICAL}/+ (n=110)	0%	3.6%	7.2%	7.3%
Sema1a^{P1}/+ (n=110)	0%	2.7%	8.1%	9.1%
Df(4)C3^{PlexA}/+ (n=100)	0%	1.0%	3.0%	12.0%
LOSS OF FUNCTION:				
Df(3R)swp2^{MICAL}/Df(3R)swp2^{MICAL}(n=103)	1.0%	68.9%	57.3%	81.2%
Sema1a^{P1}/Sema1a^{P1} (n=97)	5.2%	47.4%	77.3%	85.7%
Df(4)C3^{PlexA}/Df(4)C3^{PlexA} (n=148)	4.1%	60.8%	47.3%	74.3%
Elav-GAL4/UASMICAL; Df(3R)swp2^{MICAL}/Df(3R)swp2^{MICAL}(n=138)	0%	5.8%	13.0%	20.3%
UASMICAL$^{G\rightarrow W}$/+;Elav-GAL4/+; Df(3R)swp2^{MICAL}/Df(3R)swp2^{MICAL}(n=137)	0%	67.2%	68.6%	69.7%
GENETIC INTERACTIONS:				
Sema1a^{P1}/+;Df(3R)swp2^{MICAL}/+ (n=110)	3.6%	32.7%	37.3%	51.8%
Df(3R)swp2^{MICAL}/+;Df(4)C3^{PlexA}/+ (n=105)	0%	41.9%	33.0%	44.6%
Sema1a^{P1}/+;Df(4)C3^{PlexA}/+ (n=108)	0%	32.4%	39.8%	68.5%
Sema1a^{P1},Elav-GAL4/+; UASMICAL/Df(3R)swp2^{MICAL} (n=99)	0%	6.1%	6.1%	9.2%
GAIN OF FUNCTION:				
Elav-GAL4/Elav-GAL4; UASMICAL$^{Myr\text{-}CT}$/UASMICAL$^{Myr\text{-}CT}$(n=97)	2.1% (50%)	32.0% (3.2%)	48.5%	46.4%(2.2%)
UASMICAL/UASMICAL; Elav-GAL4/Elav-GAL4 (n=130)	0.8% (100%)	44.2% (59.6%)	50.0%	38.0% (71.4%)
UASMICAL$^{G\rightarrow W}$/+; Elav-GAL4/+ (n=170)	2.4% (100%)	36.7% (43.5%)	44.4%	48.5% (37.8%)

Description of phenotypes: [a]failure of all ISNb axons to defasciculate from the ISN; [b]ISNb axons stalling, bypassing targets, absent or decreased muscle innervation; [c]failure to make the two characteristic turns between muscles 22 and 23 and muscles 23 and 24; [x]all ISNb axons follow the ISNd or SNc, or ISNb axons remain fasciculated with the ISN but ultimately wander in the lateral muscle fields or project back to innervate ventral muscles; [y]increased (long or excessively thick) muscle innervation, excessive branching, projecting into abnormal target fields; [z]fasciculation with the ISN, premature branching, following abnormal pathways, termination on wrong muscles. (x,y,z indicate percent (%) of defects in a,b,and c, respectively).

EXAMPLE 4

MICAL GENETICALLY INTERACTS WITH SEMA1A AND PLEXA

[0388] This example illustrates that MICAL and PlexA function in the same signaling pathway to guide motor axons.

RESULTS

[0389] To address whether MICAL functions in the same signaling pathway with PlexA to mediate Sema-1a-mediated repulsive axon guidance classical genetic interaction analysis was employed by asking whether heterozygosity at both MICAL and PlexA, or MICAL and Sema1a, resulted in phenotypes not observed in either heterozygote alone. MICAL, PlexA, or Sema1a heterozygous embryos show no motor axon guidance defects (Table 1). Embryos heterozygous for both Sema1a and PlexA (Sema1a/+;Df(4)C3P^{lexA}/+) show highly penetrant axon guidance defects similar to those observed in homozygous Sema1a or PlexA mutants (Winberg et al., 1998b; Table 1). Embryos heterozygous for both MICAL and Sema1a, or heterozygous for both MICAL and PlexA, exhibit axon guidance phenotypes similar to those seen in Sema1a/+;Df(4)C3^{PlexA}/+ embryos, and these are seen at approximately equal penetrance (Table 1). For example, the ISNb and SNa of Sema1a/+; Df(3R)swp2^{MICAL}/+ or Df(3R)swp2^{MICAL}/+;Df(4)C3^{PlexA}/+ embryos exhibit guidance errors at specific choice points similar to those seen in homozygous PlexA, Sema1a, or MICAL mutant embryos. One copy each of both the UAS-MICAL and Elav-GAL4 transgenes was introduced into the Sema1a/+; Df(3R)swp2^{MICAL}/+ background and observed that neuronal MICAL expression rescues both the ISNb and SNa phenotypes in these transheterozygous embryos (Table 1). These results support the idea that MICAL and PlexA function in the same signaling pathway to guide motor axons.

EXAMPLE 5

MICAL GAIN-OF-FUNCTION AXON GUIDANCE PHENOTYPES

[0390] This example provides results that further establish that MICAL participates in PlexA-mediated motor neuron guidance, and illustrates that dominant negative MICAL mutants can be generated.

[0391] The $MICAL^{Myr\text{-}CT}$ construct was constructed by PCR amplification of the plexin interacting region of MICAL with PCR primers containing a myristoylation sequence (base pairs corresponding to the first 14 amino acids of *Drosophila* src; Simon et al, 1985), cloned into the pUAST vector, sequenced on both strands, and one *Drosophila* transgenic was obtained as described above. Embryos expressing 1 copy of $\mathit{UASMICAL}^{Myr\text{-}CT}$ in all neurons using the GAL4-UAS system (*Elav-GAL4*) exhibited phenotypes less penetrant then when 2 copies were expressed .

RESULTS

[0392] To complement MICAL LOF analysis, it was determined whether MICAL GOF mutants exhibit motor axon guidance phenotypes similar to those observed in PlexA GOF mutants (Winberg et al., 1998b). MICAL was overexpressed in all neurons in a wild type background using the GAL4-UAS system and our rescue construct. Neuronal overexpression using one or two copies of our MICAL rescue construct leads to highly penetrant motor axon guidance phenotypes (Table 1). GOF phenotypes resulting from one copy of the MICAL rescue construct in a wild type background can be suppressed in a $Df(3R)swp2^{MICAL}$ genetic background (Table 1). These defects in some cases are quite similar to the defects observed in MICAL mutants and defects reported in PlexA GOF mutants (Winberg et al., 1998b). However, a large fraction of these MICAL GOF motor axon guidance phenotypes are consistent with increased defasciculation (Table 1), as similarly described for the PlexA GOF mutants. For example, ISNb axons were often seen to abnormally leave the ISNb and project incorrectly within the ventral musculature (Table 1). Likewise, some SNa axons defasciculated at incorrect locations and projected to

inappropriate areas (Table 1). Therefore, MICAL GOF mutants exhibit phenotypes similar to PlexA GOF mutants, again suggesting that MICAL participates in PlexA-mediated motor axon guidance.

[0393] Additional support for MICAL's role in PlexA signaling was obtained by expressing in all neurons a truncated MICAL protein consisting only of the MICAL PlexA-interacting region (Figure 1B). This protein was targeted to the cell membrane by introducing an N-terminal myristoylation sequence (MICALMyr-CT) and found that neuronally expressed MICALMyr-CT acts in a dominant-negative fashion, resulting in axon guidance phenotypes similar to those observed in MICAL mutants (Table 1). Prominent GOF phenotypes like those resulting from MICAL or PlexA overexpression were not observed, indicating that neuronal MICALMyr-CT is likely occluding normal MICAL-PlexA associations and therefore MICAL signaling. This also suggests that the MICAL protein contains domains distinct from the PlexA-interacting domain that function to regulate axonal guidance.

EXAMPLE 6

THE MICALS ARE A FAMILY OF NEURONALLY EXPRESSED, PLEXIN-INTERACTING PROTEINS CONSERVED FROM FLIES TO MAMMALS

[0394] This example illustrates that MICAL proteins have conserved protein domains with identical organization in all family members.

RESULTS

[0395] MICAL proteins have conserved protein domains with identical organization in all family members and a high degree of amino acid identity among these domains in different MICALs (Figure 2A). Suzuki et al. (2002) identified MICAL-1 and a partial sequence of MICAL-2. One MICAL was identified in *Drosophila* and three mammalian MICALs were identified. The MICALs appear unique with respect to containing both calponin homology (CH) and LIM domains, in addition to their conserved N- and C-

terminal regions. A family of MICAL-like (MICAL-L) proteins were also identified, members of which have a similar organization to MICALs but lack the region N-terminal to the CH domain (Figure 2B). There is one MICAL-L protein in *Drosophila* (D-MICAL-L) and at least two family members in humans. D-MICAL-L cDNA and genomic DNA sequence information suggest that D-MICAL-L begins just N-terminal to the CH domain. Analysis of publicly available mammalian cDNA and genomic sequences suggests that human MICAL-L1 and MICAL-L2 are similar in overall domain organization to D-MICAL-L and do not contain the highly conserved ~500 amino acid MICAL N-terminal domain.

[0396] To address whether the function of MICALs is conserved in vertebrates, expression patterns and interactions with plexins were analyzed. It was found that the mRNA of all three rat MICALs shows specific neuronal and non-neuronal expression during development. For example, MICAL1, MICAL2, and MICAL3 are expressed in the rat spinal cord, dorsal root ganglia (DRG), and sympathetic ganglia at embryonic days 15 (E15) and E18 in patterns which appear overlapping but distinct. Interestingly, the neuronal expression patterns of individual MICALs are similar to those observed for several plexins, as can be seen for PlexA3 and MICAL1. In addition, results presented herein indicate that the plexin interacting domains of human MICAL-1 and mouse MICAL-2 specifically interact with the C2 domains of human PlexA3 and mouse PlexA4, respectively, and do so as strongly as the autologous domains of *Drosophila* MICAL and PlexA (see Figure 5B).

EXAMPLE 7

MICALS CONTAIN AN N-TERMINAL FLAVOPROTEIN MONOOXYGENASE DOMAIN

[0397] This example illustrates that MICAL proteins include a highly conserved an N-terminal monooxygenases domain.

MICAL Flavoprotein Monooxygenase Fusion Protein Purification and FAD Binding

[0398] A His-tagged bacterial fusion protein was constructed that included the MICAL flavoprotein monooxygenase domain (MICAL FM) by inserting amino acids 1-526 of *Drosophila* MICAL into the bacterial expression vector pET 43.1b containing a hexahistidine tag (Novagen). The plasmid was transformed into E. Coli BL21 (DE3) and the hexahistidine tagged recombinant protein was expressed by IPTG induction and MICAL FM was isolated with the inclusion bodies, denatured with 6M GdmHCl (Gibco) and purified under denaturing conditions over a Ni^{2+} column (Novagen). MICAL FM was renatured at 25°C for 3 hours by diluting the purified protein 100X into a solution containing a 5-fold molar excess (to MICAL FM present in the solution) of free FAD (Sigma), 10mM DTT, and 10μg/mL BSA as described (Lindsay et al., 2000). Renatured protein was dialyzed for 48 hours at 4°C into 5 changes of His binding buffer (5mM imidazole, 0.5 M NaCl, 20mM Tris-HCl, pH 7.9) to remove free FAD, and DTT. Ni^2+ purification beads were then incubated with the MICAL FM sample in batch for 5 hours. at 4°C. The solution was then repurified through a Ni^{2+} column. Fractions containing MICAL FM were pooled and subjected to dialysis into a more stable buffer containing 5mM DTT, and then subjected to Coomassie staining and Western analysis to confirm the purity of the sample. Spectral analysis was done using a Perkin-Elmer UV/VIS Lambda-12 spectrophotometer scanning from 300 to 550nm.

RESULTS

[0399] The high degree of conservation of the MICAL N-terminus among family members (up to 62% identical between flies and humans; Figure 2A) suggests that this domain is functionally important. Upon closer examination of this conserved region, we noted a consensus dinucleotide binding sequence, GXGXXG (Figures 1B and 3A), which is distinct from the sequence present in classical mononucleotide binding motifs (Eggink et al., 1990; Eppink et al., 1997; Schulz, 1992; Wierenga et al., 1986). Further, the amino acid sequence in this 500 amino acid region reveals that MICALs contain three separate sequence motifs spaced throughout this domain that define them as flavoprotein

monooxygenases (also called hydroxylases), a subclass of oxidoreductases (Eggink et al., 1990; Eppink et al., 1997; Wierenga et al., 1986). The amino acid sequence surrounding the GXGXXG motif matches perfectly the consensus sequence for the ADP binding region of flavin adenine dinucleotide (FAD) binding proteins (Rossmann fold or FAD Fingerprint 1, Figures 1B and 3A), and distinguishes this region from consensus NAD, or NADP binding folds (Vallon, 2000; Wierenga et al., 1986). MICALs also have a well-conserved GD motif (FAD Fingerprint 2; Figures 1B and 3A) C-terminal to the FAD Fingerprint 1 region, which is important for binding the ribose moiety of FAD (Eggink et al., 1990; Eppink et al., 1997). Finally, MICALs have the conserved DG motif ("Conserved Motif"; Figures 1B and 3A) between the FAD Fingerprint 1 and 2 motifs that has been reported to be involved in binding the pyrophospate moiety of FAD (Eppink et al., 1997). Proteins with these consensus FAD binding regions bind FAD and use FAD in the catalysis of oxidation-reduction reactions. Flavoprotein monooxygenases are oxidoreductases (enzymes that catalyze oxidation and reduction reactions) and catalyze the insertion of one atom of molecular oxygen into their substrate using nucleotides as electron donors (Massey, 1995). These monooxygenases are also defined by their use of FAD as a co-enzyme. Apart from these three consensus regions, monooxygenases vary significantly, reflecting the wide range of enzymes in this family and their variable substrate binding pockets also encompassed within this domain (Eppink et al., 1997). However, MICALs and other monooxygenases show significant similarity within these three FAD binding regions and also similar spacing of these regions within the monooxygenase domain.

[0400] Does MICAL bind FAD? A solution of the purified MICAL-flavoprotein monooxygenase (FM) domain (expressed in bacteria) is yellow in color, a characteristic of flavoproteins. Spectral analysis of purified MICAL-FM shows that it has an absorption peak at 452nm and a shoulder at ~358nm (Figure 3B). This is similar to the absorption spectra of FAD itself (~450nm and ~360nm; Macheroux, 1999), and to other related flavoproteins (e.g., p-Hydroxybenzoate Hydroxylase, Hosokawa and Stanier, 1966; and GidA, White et al., 2001), suggesting that MICAL-FM binds FAD. These results, in combination with the sequence homology, raises the possibility that MICAL enzymatic activity within the N-terminal conserved domain serves an integral function in plexin signaling.

EXAMPLE 8

AN INTACT FAD BINDING MOTIF IS REQUIRED FOR MICAL MOTOR AXON GUIDANCE FUNCTIONS

[0401] This example illustrates that an intact flavoprotein monooxygenases domain is necessary for MICAL function in repulsive motor axon guidance.

[0402] To make the MICAL$^{G\rightarrow W}$ mutant (SEQ ID NO:20), the dinucleotide binding region of *Drosophila* MICAL was mutated from GXGXXG to WXWXXW, such that the glycines were changed to tryptophans. Oligonucleotides containing terminal endogenous restriction sites (Mlu I) and base pair substitutions (<u>G</u>G<u>A</u> GCA <u>G</u>GG CCC TGT <u>G</u>G<u>A</u> (SEQ ID NO:37) changed to <u>T</u>G<u>G</u> GCA <u>T</u>GG CCC TGT <u>T</u>G<u>G</u> (SEQ ID NO:38)) were used to amplify a 1.4kb fragment that was cloned in the correct orientation into the full length MICAL rescue construct. The region was sequenced on both stands and substitutions also disrupted a restriction site (Apa I) so the mutated construct could also be confirmed by restriction analysis. One transgenic, located on the X chromosome, was obtained.

RESULTS

[0403] The glycine residues within the GXGXXG motif of FAD binding proteins are essential for allowing the FAD binding and enzymatic activity (Wierenga et al, 1986; Dym and Eisenberg, 2001). To test the necessity of MICAL FAD binding, and potential enzymatic activity, for plexin-mediated repulsive axon guidance the three glycine residues within the FAD Fingerprint 1 motif of MICAL were mutated to tryptophan (<u>G</u>A<u>G</u>PC<u>G</u>L (SEQ ID NO:39)→<u>W</u>A<u>W</u>PC<u>W</u>L (SEQ ID NO:40): mutations which in related flavin-containing monooxygenases disrupt FAD binding but do not alter the overall structure of the protein (Kubo et al., 1997; Lawton and Philpot, 1993; Wierenga et al., 1986). The resulting construct, MICAL$^{G\rightarrow W}$, was used for *in vivo* neuronal expression in *Drosophila*. Transgenic flies containing the UAS-MICAL$^{G\rightarrow W}$ transgene were generated and immunohistochemical and Western analyses confirmed that MICAL$^{G\rightarrow W}$ was expressed at levels comparable to those of our wild-type "short" MICAL variant that was used to

rescue MICAL mutant motor axon guidance phenotypes . Unlike neuronal expression of MICAL in a homozygous Df(3R)swp2MICAL mutant background, which rescues all ISNb and SNa defects, one copy of the neuronal $MICAL^{G\rightarrow W}$ rescues none of these defects (Table 1). This strongly suggests that activity of the MICAL monooxygenase domain is necessary for normal MICAL function.

[0404] Since $MICAL^{G\rightarrow W}$ contains an intact plexin interacting domain but is functionally inactive, we predicted that it would exert a dominant-negative effect on motor axon projections in a wild-type genetic background, binding to PlexA but blocking signaling in a manner similar to the MICALMyr-CT construct. When one copy of the $MICAL^{G\rightarrow W}$ reporter construct was used to express $MICAL^{G\rightarrow W}$ in all neurons in a wild-type genetic background we observed highly penetrant ISNb, SNa, and CNS longitudinal connective defects (Table 1) providing further evidence that $MICAL^{G\rightarrow W}$ is being expressed neuronally and is likely able to bind PlexA. However, though many of these defects resemble phenotypes observed in Sema1a, PlexA, or MICAL LOF mutants (Table 1), a significant fraction (ISNb: >44%, SNa: 38%) were strikingly distinct (Table 1). For example, though it was observed that ISNb and SNa axon guidance phenotypes consistent with MICAL LOF phenotypes (Table 1), these phenotypes were often more severe. They include defects in which axons bypass their muscle targets but then appear to defasciculate in inappropriate places and project into adjacent segments. Interestingly, also observed were ISNb and SNa axon guidance phenotypes consistent with MICAL GOF, but these phenotypes also appeared more severe and included, severely defasciculated and tangled axons. Finally, phenotypes were observed that were unlike MICAL or PlexA LOF or GOF mutants, including axons projecting along the entire length of the muscle 6/7 cleft and dramatic axonal wandering within muscle fields. These phenotypes suggest that expression of $MICAL^{G\rightarrow W}$ leads to defects not explained by a simple elevation or diminution of PlexA signaling activity.

[0405] In summary, these results show the necessity of an intact flavoprotein monooxygenase domain for MICAL function in repulsive motor axon guidance.

EXAMPLE 9

FLAVOPROTEIN MONOOXYGENASE INHIBITORS NEUTRALIZE VERTEBRATE SEMA3A AXONAL REPULSION

[0406] This example identifies the gallic acid derivatives EGCG and EC as inhibitors of semaphorin-mediated axon repulsion.

Vertebrate *In vitro* Repulsion and Collapse Assays

[0407] DRG repulsion assays were performed as described (Messersmith et al., 1995). (EGCG), (EC), L-NAME, allopurinol, (Sigma) were dissolved in vehicle (PBS), protected from light, and then added to the culture media to final concentrations. Rotenone (Sigma) was dissolved in 95% EtOH and then added to the culture media (final EtOH concentration was below 0.1% and had no effect on axon outgrowth).

[0408] Inhibitors specific for nitric oxide synthase (N-nitro-L-arginine methylester (L-NAME); Comoletti et al., 2001), xanthine oxidase (allopurinol (Allo); Jonakait et al., 2000), or mitochondrial electron transport (NADH dehydrogenase; rotenone (Rote); Frantseva et al., 2001) were used at concentrations previously shown to be effective in cell culture conditions (see Figure 7). The effect of EGCG on growth cone collapse was performed by culturing DRGs using standards techniques (Fan et al, 1993). After 48 hours in culture, DRGs were incubated in media containing EGCG or vehicle (PBS) for 3 hours prior to a 1 hour application of 1nM AP-Sema3A or non-AP-Sema3A containing growth media. Scoring was done by splitting explants into quadrants and scoring all growth cones as either collapsed or not-collapsed.

RESULTS

[0409] The MICALs may be susceptible to small molecule inhibitors that affect their ability to oxidize their substrate. Some gallic acid derivatives, including the green tea

component (-)-epigallocatechin gallate (EGCG), are potent and selective inhibitors of two flavoprotein monooxygenases: squalene epoxidase (SE) and p-hydroxybenzoate hydroxylase (pHBH) (Abe et al., 2000a; Abe et al., 2000b).

[0410] All available evidence points to the plexin cytoplasmic domain as an essential signal transducing domain for signaling class 3 semaphorin repulsion (Cheng et al., 2001; Takahashi and Strittmatter, 2001). Sema3A appears to utilize neuropilin-1 in combination with A class plexins to signal repulsive guidance. To ask whether selective flavoprotein monooxygenase inhibitors can neutralize semaphorin-mediated repulsion in vertebrates, *in vitro* rat DRG growth cone repulsion assays were employed using Sema 3A-secreting 293 cells (Figure 7A; Messersmith et al., 1995). NGF-dependent DRG axons exhibit little to no outgrowth towards Sema3A-secreting 293 cell aggregates (Figures 4C). However, when 25 mM EGCG is added to the growth media Sema3A repulsion was completely neutralized (Figures 4C). EGCG attenuation of Sema3A-mediated repulsion is dose-dependent (Figure 4C). We also asked whether (-)-epicatechin (EC), a compound structurally similar to EGCG but a poor inhibitor of SE (Abe et al., 2000b), had a similar effect on Sema-3A-mediated repulsion. Like EGCG, EC was capable of completely neutralizing Sema-3A-dependent repulsion in a dose-dependent manner, but a much higher EC concentration was required (Figures 4C). To address the possibility that a general inhibition of oxidation-reduction mechanisms by these reagents underlies this attenuation of Sema3A repulsion, selective inhibitors of other redox enzymes present in neurons were analyzed for an effect on Sema3A-mediated repulsion. No attenuation of Sema 3A mediated axonal repulsion was observed using inhibitors specific for nitric oxide synthase (N-nitro-L-arginine methylester (L-NAME)), xanthine oxidase (allopurinol (Allo), or mitochondrial electron transport (NADH dehydrogenase; rotenone (Rote), at concentrations previously shown to be effective in cell culture conditions (Figure 4C). DRG axons and Sema3A-secreting 293 cells appeared normal following growth in the presence of all but one of these inhibitors. In some explants we noticed an adverse effect on survival of DRGs treated with rotenone, but axons in those rotenone-treated explants that survived, although somewhat thinner than normal, were robustly repelled (Figure 4C). The amount and biological activity of Sema3A produced by 293 cells in the presence of all inhibitors was similar as assessed using a DRG growth cone collapse assay (Figure 4B), showing that none of these inhibitors had an

adverse effect on the ability of 293 cells to produce active Sema 3A. It was also determined in separate experiments that 25mM EGCG dramatically abrogates Sema3A-mediated growth cone collapse in NGF-dependent DRG neurons. Taken together, our results support a role for flavoenzymes and oxidation-reduction mechanisms in semaphorin-mediated axon guidance.

EXAMPLE 10

FURTHER CONSIDERATIONS REGARDING THE ASSOCIATION OF MICALS AND SEMAPHORIN-MEDIATED AXONAL GUIDANCE

[0411] This example provides further insight into the association of MICALs and semaphorin-mediated axonal guidance.

[0412] Neuronal growth cone guidance depends on the ability of various guidance cue receptors to regulate cytoskeletal dynamics in response to the local presentation of ligands. It was shown herein that proteins belonging to the MICAL family of cytosolic, multi-domain, flavoprotein monooxygenases are required for certain plexin-mediated semaphorin axon guidance events. MICALs associate with plexins and contain several conserved domains that provide the potential for interactions with both growth cone cytoskeletal components and many signaling proteins intimately involved in their regulation. Our results suggest that MICALs directly participate in plexin signaling through the action of their flavoprotein monooxygenase domain. These observations provide a framework for dissecting the molecular basis of semaphorin-meditated neuronal guidance and also a potential target for attenuating their repulsive action.

[0413] Genetic and biochemical results provided herein support an essential role for *Drosophila* MICAL in mediating PlexA/Sema-1a repulsive guidance events required for motor axon pathfinding. Future experiments will establish whether MICAL mediates PlexB signaling, and if so, whether this occurs directly or indirectly. *Drosophila* MICAL is an orthologue of a mammalian MICAL-1 protein. It is shown herein that that there are at least three vertebrate MICAL orthologues (MICALs 1, 2, and 3). We also identify here a family of MICAL-like proteins that lack the conserved N-terminal MICAL monooxygenase

domain. Expression and interaction data herein support the idea that MICALs mediate plexin signaling in vertebrates. In addition, flavoprotein monooxygenase inhibitors block Sema3A-mediated repulsion and collapse of NGF-dependent DRG axons—repulsive interactions dependent on A class plexins including Plexin A3. Future genetic and biochemical analysis will establish the role of vertebrate MICALs in neuronal and non-neuronal plexin signaling.

[0414] The highly conserved ~500 amino acid N-terminal MICAL domain contains signature amino acid sequences of the flavoprotein monooxygenase family of oxidoreductases. Biochemical and genetic analyses herein strongly suggest that MICALs contain functional FAD binding monooxygenase domains required for mediating plexin signaling. In support of this idea, it was observed that inhibition of flavoprotein monooxygenase enzymatic activity dramatically attenuates semaphorin-mediated axon repulsion and growth cone collapse. However, though the inhibitors we used, ECGC and EC, have a high degree of selectivity for flavoprotein monooxygenases, similar concentrations of EGCG inhibit other enzymes including steroid 5a-reductase, NADPH-cytochrome P450 reductase, telomerase, matrix metalloproteinases MMP-2 and MMP-9, and phenol sulfotransferase (Abe et al., 2000a; Abe et al., 2000b). Although most of these enzymes are unlikely to be expressed in the growth cones of DRG axons, potential non-specific effects of these inhibitors cannot be ruled out despite their demonstrated selectivity for monooxygenases. Taken together with our *in vivo Drosophila* experiments showing a requirement for the MICAL FAD binding region in Sema-1a mediated axon repulsion, these data suggest redox signaling plays an important role in vertebrate semaphorin-mediated axonal repulsion.

[0415] Flavoprotein monooxygenases specifically catalyze the oxidation of a number of substrates, and in some contexts they can function as oxidases and generate reactive oxygen species (Massey, 1994). Results herein suggest that MICALs are flavoproteins most similar to the flavoprotein monooxygenase family of oxidoreductases, but a complete understanding of the chemical nature of the reactions catalyzed by MICALs awaits future study and identification of substrates. The redox regulation of amino acid residues within signaling proteins (including kinases, phosphatases, small GTPases, guanylate cyclases, and

adapter proteins) and cytoskeletal proteins (including actin, actin binding proteins, intermediate filament proteins, and GAP-43) has been shown to modulate their function (Finkel, 1998; Kim et al., 2002; Meng et al., 2002; Rhee et al., 2000; Stamler et al., 2001; Thannickal and Fanburg, 2000). In addition, oxidation of actin leads to disassembly of actin filaments, instability and collapse of actin networks, reduced ability of actin to interact with actin cross-linking proteins, and a decrease in the ability of actin monomers to form polymers (Dalle-Donne et al., 2001a; Dalle-Donne et al., 2001b; Milzani et al., 1997). Finally, it is also interesting that MICALs have a putative actin filament binding domain (CH domain) and that MICAL-1 interacts with vimentin, an intermediate filament protein (Suzuki et al., 2002).

[0416] It was recently reported that the proline rich region of vertebrate MICAL-1 interacts with the SH3 domain of the adaptor protein CasL (HEF1) in non-neuronal cells (Suzuki et al., 2002). CasL, along with the related proteins p130Cas, and Efs (Sin), make up the Cas family of proteins (O'Neill et al., 2000), which assemble and transduce intracellular signals that stimulate cell migration and axon outgrowth. These proteins have numerous protein-protein interaction domains, including a Src-homology 3 (SH3) domain, multiple SH2-binding sites in their substrate domain, several proline-rich motifs, and a C-terminal dimerization module. This structure suggests a role for Cas family proteins as docking molecules, and numerous interacting proteins have been identified, including kinases (e.g. FAK, Src, and Abl), phosphatases (e.g. PTP-1B, and SHP2), GEFs (e.g. C3G), and adaptor proteins (e.g., Nck, Crk, Grb2, and 14-3-3) (O'Neill et al., 2000). Studies indicate that Cas proteins localize mainly to focal adhesions and stress fibers, and that they are required in integrin-dependent cell migration and actin filament assembly. Cas proteins, therefore, may play an important role in plexin-mediated repulsive and attractive guidance events.

[0417] In conclusion, characterized herein is a gene family conserved from invertebrates to vertebrates, with proteins whose structure and function strongly suggest that redox signaling is important for semaphorin-mediated axonal repulsion. The results herein also suggest that protein oxidation could be a general means for inhibiting axonal growth. Given the presence of high amounts of reactive oxygen species and other oxidants in the spinal

cord after injury (Juurlink and Paterson, 1998) regulation of redox signaling using antioxidants and specific enzyme inhibitors may be a powerful approach for encouraging axonal regeneration.

Table 2. List of sequences

SEQ ID NO:	Sequence
1	Human MICAL 1 cDNA
2	Human MICAL 1 polypeptide
3	Human MICAL 2 cDNA
4	Human MICAL 2 polypeptide
5	Human MICAL 3 cDNA
6	Human MICAL 3 polypeptide
7	Drosophila MICAL long variant cDNA
8	Drosophila polypeptide (long variant)
9	Drosophila MICAL medium variant cDNA
10	Drosophila polypeptide (medium variant)
11	Drosophila short variant cDNA
12	Drosophila polypeptide (short variant)
13	Human MICAL-Like 1 cDNA
14	Human MICAL-Like 1 polypeptide
15	Human MICAL-Like 2 cDNA
16	Human MICAL-Like 2 polypeptide
17	Drosophila MICAL-Like cDNA
18	Drosophila MICAL-Like polypeptide
19	Drosophila truncated mutant polypeptide
20	Drosophila G to W mutant polypeptide
21	Mouse MICAL 1 polypeptide
22	Mouse MICAL 2 polypeptide
23	Mouse MICAL 3 polypeptide
24	*Anopheles gambiae* MICAL polypeptide fragment
25	*Ciona inetstinalis* MICAL polypeptide fragment
26	*Danio rerio* MICAL 1 polypeptide fragment
27	*Danio rerio* MICAL 2 polypeptide fragment
28	*Gallus gallus* MICAL 1 polypeptide fragment
29	*Gallus gallus* MICAL 2 polypeptide fragment
30	*Rattus norvegicus* MICAL 1 polypeptide fragment
31	*Rattus norvegicus* MICAL 2 polypeptide fragment
32	*Rattus norvegicus* MICAL 3 polypeptide fragment
33	*Bos taurus* MICAL 1 polypeptide fragment
34	*Bos taurus* MICAL 2 polypeptide fragment
35	*Sus scrofa* MICAL polypeptide fragment
36	*Pan troglodytes* MICAL polypeptide fragment
37	Amplification primer for MICAL
38	Amplification primer for mutant MICAL
39	FAD binding domain
40	mutated FAD binding domain

References

[0418] Abe, I., Kashiwagi, K., and Noguchi, H. (2000a). Antioxidative galloyl esters as enzyme inhibitors of p-hydroxybenzoate hydroxylase, *FEBS Lett.* 483, 131-4.

[0419] Abe, I., Seki, T., Umehara, K., Miyase, T., Noguchi, H., Sakakibara, J., and Ono, T. (2000b). Green tea polyphenols: novel and potent inhibitors of squalene epoxidase. *Biochem. Biophys. Res. Commun.* 268, 767-71.

[0420] Bach, I. (2000). The LIM domain: regulation by association. *Mech. Dev.* 91, 5-17.

[0421] Brand, A. H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 118, 401-415.

[0422] Bretscher, A., Chambers, D., Nguyen, R., and Reczek, D. (2000). ERM-Merlin and EBP50 protein families in plasma membrane organization and function. *Annu. Rev. Cell Dev. Biol.* 16, 113-43.

[0423] Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001). Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* 11, 82-8.

[0424] Cheng, H. J., Bagri, A., Yaron, A., Stein, E., Pleasure, S. J., and Tessier-Lavigne, M. (2001). Plexin-A3 mediates semaphorin signaling and regulates the development of hippocampal axonal projections. *Neuron.* 32, 249-63.

[0425] Comoletti, D., Muzio, V., Capobianco, A., Ravizza, T., and Mennini, T. (2001). Nitric oxide produced by non-motoneuron cells enhances rat embryonic motoneuron sensitivity to excitotoxins: comparison in mixed neuron/glia or purified cultures. *J Neurol Sci.* 192, 61-9.

[0426] Cook, K., Coulson, D., Roote, J., Morley, T., Bogart, K., Deal, J., Kaufman, T., and Gubb, D. (2001). Chromosomal deletions and inversions with well-defined endpoints can be generated at high frequencies by transposase-induced recombination between pairs of P-element insertions. *Drosophila.* Research Conference *42*, 19W.

[0427] Cooley, L., Thompson, D., and Spradling, A. C. (1990). Constructing deletions with defined endpoints in *Drosophila*. Proc Natl Acad Sci U S A *87*, 3170-3.

[0428] Dalle-Donne, I., Rossi, R., Giustarini, D., Gagliano, N., Lusini, L., Milzani, A., Di Simplicio, P., and Colombo, R. (2001a). Actin carbonylation: from a simple marker of protein oxidation to relevant signs of severe functional impairment. *Free Radic Biol Med.* 31, 1075-83.

[0429] Dalle-Donne, I., Rossi, R., Milzani, A., Di Simplicio, P., and Colombo, R. (2001b). The actin cytoskeleton response to oxidants: from small heat shock protein phosphorylation to changes in the redox state of actin itself. *Free Radic Biol Med.* 31, 1624-32.

[0430] Dym, O., and Eisenberg, D. (2001). Sequence-structure analysis of FAD-containing proteins. *Protein Sci.* 10, 1712-28.

[0431] Eggink, G., Engel, H., Vriend, G., Terpstra, P., and Witholt, B. (1990). Rubredoxin reductase of Pseudomonas oleovorans. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J Mol Biol.* 212, 135-42.

[0432] Golemis, E. A., Gyuris, J., and Brent, R. (1994). Interaction trap/two hybrid system to identify interacting proteins. *In Current Protocols in Molecular Biology.* (New York:Wiley) 27, pp. 13.14.1-13.14.17.

[0433] Eppink, M. H., Schreuder, H. A., and Van Berkel, W. J. (1997). Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci.* 6, 2454-8.

[0434] Fan, J., Mansfield, S. G., Redman, T., Phillip, R., Gordon-Weeks, P. R., and Raper, J. A. (1993). The organization of F-actin and microtubules in growth cones exposed to a brain-derived collapsing factor. *J Cell Biol.* 121, 867-878.

[0435] Finkel, T. (1998). Oxygen radicals and signaling. *Curr Opin Cell Bio.l* 10, 248-53.

[0436] Frantseva, M. V., Carlen, P. L., and Perez Velazquez, J. L. (2001). Dynamics of intracellular calcium and free radical production during ischemia in pyramidal neurons. *Free Radic Biol Med.* 31, 1216-27.

[0437] Gimona, M., Djinovic-Carugo, K., Kranewitter, W. J., and Winder, S. J. (2002). Functional plasticity of CH domains. *FEBS Lett.* 513, 98-106.

[0438] Harris, B. Z., and Lim, W. A. (2001). Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci.* 114, 3219-31.

[0439] He, Z., Wang, K. C., Koprivica, V., Ming, G., and Song, H. J. (2002). Knowing how to navigate: mechanisms of semaphorin signaling in the nervous system. *Sci STKE.* 2002, RE1.

[0440] Hosokawa, K., and Stanier, R. Y. (1966). Crystallization and properties of p-hydroxybenzoate hydroxylase from Pseudomonas putida. *J Biol Chem.* 241, 2453-60.

[0441] Hu, H., Marton, T. F., and Goodman, C. S. (2001). Plexin B mediates axon guidance in *Drosophila* by simultaneously inhibiting active Rac and enhancing RhoA signaling. *Neuron.* 32, 39-51.

[0442] Jonakait, G. M., Wen, Y., Wan, Y., and Ni, L. (2000). Macrophage cell-conditioned medium promotes cholinergic differentiation of undifferentiated progenitors and synergizes with nerve growth factor action in the developing basal forebrain. *Exp Neurol.* 161, 285-96.

[0443] Juurlink, B. H., and Paterson, P. G. (1998). Review of oxidative stress in brain and spinal cord injury: suggestions for pharmacological and nutritional management strategies. *J Spinal Cord Med.* 21, 309-34.

[0444] Kim, S. O., Merchant, K., Nudelman, R., Beyer, J., W.F., Keng, T., DeAngelo, J., Hausladen, A., and Stamler, J. S. (2002). OxyR: A molecular code for redox-related signaling. *Cell.* 109, 383-396.

[0445] Kolodkin, A. L., Matthes, D., and Goodman, C. S. (1993). The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell.* 75, 1389-1399.

[0446] Kubo, A., Itoh, S., Itoh, K., and Kamataki, T. (1997). Determination of FAD-binding domain in flavin-containing monooxygenase 1 (FMO1). *Arch Biochem Biophys.* 345, 271-7.

[0447] Landgraf, M., Bossing, T., Technau, G. M., and Bate, M. (1997). The origin, location, and projections of the embryonic abdominal motorneurons of *Drosophila*. *J Neurosci.* 17, 9642-55.

[0448] Lawton, M. P., and Philpot, R. M. (1993). Functional characterization of flavin-containing monooxygenase 1B1 expressed in Saccharomyces cerevisiae and Escherichia coli and analysis of proposed FAD- and membrane-binding domains. *J Biol Chem.* 268, 5728-34.

[0449] Lindsay, H., Beaumont, E., Richards, S. D., Kelly, S. M., Sanderson, S. J., Price, N. C., and Lindsay, J. G. (2000). FAD insertion is essential for attaining the assembly competence of the dihydrolipoamide dehydrogenase (E3) monomer from Escherichia coli. *J Biol Chem.* 275, 36665-70.

[0450] Liu, B. P., and Strittmatter, S. M. (2001). Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr Opin Cell Biol.* 13, 619-26.

[0451] Macheroux, P. (1999). UV-visible spectroscopy as a tool to study flavoproteins. *Methods Mol Biol.* 131, 1-7.

[0452] Massey, V. (1994). Activation of molecular oxygen by flavins and flavoproteins. *J Biol Chem.* 269, 22459-62.

[0453] Massey, V. (1995). Introduction: flavoprotein structure and mechanism. *Faseb.* J 9, 473-5.

[0454] Matthes, D. J., Sink, H., Kolodkin, A. L., and Goodman, C. S. (1995). Semaphorin II can function as a selective inhibitor of specific synaptic arborizations in *Drosophila*. *Cell.* 81, 631-639.

[0455] Meng, T. C., Fukada, T., and Tonks, N. K. (2002). Reversible oxidation and inactivation of protein tyrosine phosphatases *in vivo*. *Mol Cell.* 9, 387-99.

[0456] Messersmith, E. K., Leonardo, E. D., Shatz, C. J., Tessier-Lavigne, M., Goodman, C. S., and Kolodkin, A. L. (1995). Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord. *Neuron.* 14, 949-959.

[0457] Milzani, A., DalleDonne, I., and Colombo, R. (1997). Prolonged oxidative stress on actin. *Arch Biochem Biophys.* 339, 267-74.

[0458] O'Neill, G. M., Fashena, S. J., and Golemis, E. A. (2000). Integrin signalling: a new Cas(t) of characters enters the stage. *Trends Cell Biol.* 10, 111-9.

[0459] Pasterkamp, R. J., De Winter, F., Holtmaat, A. J. G. D., and Verhaagen, J. (1998). Evidence for a role of the chemorepellent semaphorin III and its receptor neuropilin-1 in the regeneration of primary olfactory axons. *J Neurosci.* 18, 9962-9976.

[0460] Preston, C. R., Sved, J. A., and Engels, W. R. (1996). Flanking duplications and deletions associated with P-induced male recombination in *Drosophila*. *Genetics* 144, 1623-38.

[0461] Raper, J. A. (2000). Semaphorins and their receptors in vertebrates and invertebrates. *Curr Opin Neurobiol.* 10, 88-94.

[0462] Rhee, S. G., Bae, Y. S., Lee, S. R., and Kwon, J. (2000). Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci STKE.* 2000, E1.

[0463] Schulz, G. E. (1992). Binding of nucleotides by proteins. *Curr Opin Struct Biol.* 2, 61-67.

[0464] Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Second Edition (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory).

[0465] Simon, M.A., Drees, B., Kornberg, T., and J.M. Bishop. (1985) The Nucleotide Sequence and the Tissue-Specific Expression of *Drosophila* c-src. Cell 42, 831-840.Semaphorin Nomenclature Committee (1999). Unified nomenclature for the semaphorins/collapsins. Semaphorin Nomenclature Committee, *Cell.* 97, 551-2.

[0466] Stamler, J. S., Lamas, S., and Fang, F. C. (2001). Nitrosylation. the prototypic redox-based signaling mechanism. *Cell.* 106, 675-83.

[0467] Suzuki, T., Nakamoto, T., Ogawa, S., Seo, S., Matsumura, T., Tachibana, K., Morimoto, C., and Hirai, H. (2002). MICAL, a novel CasL interacting molecule, associates with vimentin. *J Biol Chem.* 277, 14933-41.

[0468] Takahashi, T., and Strittmatter, S.M. (2001). PlexinA1 Autoinhibition by the Plexin Sema Domain. *Neuron*. 29, 429-439.

[0469] Tamagnone, L., and Comoglio, P. M. (2000). Signalling by semaphorin receptors: cell guidance and beyond. *Trends Cell Biol*. 10, 377-83.

[0470] Thannickal, V. J., and Fanburg, B. L. (2000). Reactive oxygen species in cell signaling. Am J Physiol Lung. *Cell Mol Physiol*. 279, L1005-28.

[0471] Vallon, O. (2000). New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure. *Proteins*. 38, 95-114.

[0472] VanVactor, D., Sink, H., Fambrough, D., Tsoo, R., and Goodman, C. S. (1993). Genes that control neuromuscular specificity in *Drosophila*. *Cell*. 73, 1137-1153.

[0473] White, D. J., Merod, R., Thomasson, B., and Hartzell, P. L. (2001). GidA is an FAD-binding protein involved in development of Myxococcus xanthus. *Mol Microbiol*. 42, 503-17.

[0474] Whitford, K. L., and Ghosh, A. (2001). Plexin signaling via off-track and rho family GTPases. *Neuron*. 32, 1-3.

[0475] Wierenga, R. K., Terpstra, P., and Hol, W. G. (1986). Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol*. 187, 101-7.

[0476] Winberg, M. L., Mitchell, K. J., and Goodman, C. S. (1998a). Genetic analysis of the mechanisms controlling target selection: complementary and combinatorial functions of netrins, semaphorins, and IgCAMs. *Cell*. 93, 581-591.

[0477] Winberg, M. L., Noordermeer, J. N., Tamagnone, L., Comoglio, P. M., Spriggs, M. K., Tessier-Lavigne, M., and Goodman, C. S. (1998b). Plexin A is a neuronal semaphorin receptor that controls axon guidance. *Cell*. 95, 903-916.

[0478] Xu, X. M., Fisher, D. A., Zhou, L., White, F. A., Ng, S., Snider, W. D., and Luo, Y. (2000). The transmembrane protein semaphorin 6A repels embryonic sympathetic axons. *J. Neurosci.* 20, 2638-48.

[0479] Yu, H. H., Araj, H. H., Ralls, S. A., and Kolodkin, A. L. (1998). The transmembrane Semaphorin Sema I is required in *Drosophila* for embryonic motor and CNS axon guidance. *Neuron* 20, 207-20.

[0480] Although the invention has been described with reference to the above example, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. An isolated polypeptide comprising an N-terminal MICAL domain, a calponin homology domain, a LIM domain, a proline rich region, and a plexin interacting region, wherein the polypeptide has monooxygenase activity.

2. An isolated polypeptide of claim 1, wherein the polypeptide is a mammalian MICAL polypeptide.

3. An isolated polypeptide of claim 2, wherein the isolated polypeptide is human MICAL-1, human MICAL-2, or human MICAL-3.

4. An isolated polypeptide of claim 3, wherein the polypeptide comprises an amino acid sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

5. An isolated polypeptide of claim 1, wherein the polypeptide comprises an N-terminal MICAL domain having at least about 50% sequence identity to the N-terminal amino acids 1-500 of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

6. An isolated polypeptide of claim 1, wherein the polypeptide is a *Drosophila* MICAL polypeptide.

7. An isolated polypeptide of claim 6, wherein the polypeptide is set forth in SEQ ID NO:8.

8. An isolated polypeptide of claim 1, wherein the polypeptide is a MICAL isoform.

9. An isolated polypeptide of claim 1, wherein the isolated polypeptide is at least 90% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

10. An isolated polypeptide of claim 1, wherein the polypeptide comprises from N-terminal to C-terminal, an N-terminal MICAL domain, a calponin homology domain, a first variable MICAL region, a LIM domain, a proline rich region, and a plexin interacting region.

11. An isolated polypeptide comprising a plexin interacting region at least 90% identical to a plexin interacting region of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12, wherein the polypeptide has plexin interacting activity.

12. An isolated polypeptide of claim 11, wherein the polypeptide comprises a plexin interacting region of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12, or a conservative variant thereof.

13. An isolated polypeptide of claim 11, with the proviso that the polypeptide does not have monooxygenase activity.

14. An isolated polypeptide of claim 11, wherein the polypeptide comprises the plexin interacting region of *Drosophila* MICAL-like polypeptide, or the plexin interacting region of human MICAL-like polypeptide 1, or 2, or a conservative variant thereof.

15. An isolated polypeptide of claim 14, wherein the polypeptide has the amino acid sequence of SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

16. An isolated polypeptide comprising an N-terminal MICAL domain of *Drosophila* MICAL 1, or the N-terminal MICAL domain of human MICAL 1, 2, or 3, or a conservative variant thereof.

17. An isolated polypeptide of claim 16, wherein the polypeptide has monooxygenase activity.

18. An isolated polypeptide comprising a calponin homology domain at least 90% identical to the calponin homology domain of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12, and wherein the polypeptide is involved in actin filament binding.

19. An isolated polypeptide comprising a LIM domain at least 90% identical to the LIM domain of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12, and wherein the polypeptide specifically interacts with a LIM-binding protein.

20. An isolated polypeptide comprising a proline rich region at least 90% identical to the proline rich region of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12, and wherein the polypeptide interacts with a polypeptide comprising an SH3-domain

21. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide of claim 1.

22. An isolated polynucleotide of claim 21, wherein the polynucleotide encodes a mammalian MICAL polypeptide.

23. An isolated polynucleotide of claim 22, wherein the polynucleotide encodes a polypeptide that is at least 90% identical to SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

24. An isolated polynucleotide of claim 23, wherein the polynucleotide encodes a polypeptide as set forth in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

25. An isolated polynucleotide of claim 21, wherein the polynucleotide encodes a MICAL polypeptide comprising an N-terminal MICAL domain having monooxygenase activity, and at least 50% sequence identity to the N-terminal 500 amino acids of human MICAL 1 polypeptide.

26. An isolated polynucleotide of claim 21, wherein the polynucleotide encodes a *Drosophila* MICAL polypeptide.

27. An isolated polynucleotide of claim 26, wherein the polynucleotide encodes an amino acid sequence as set forth in SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

28. A vector comprising a polynucleotide of claim 15.

29. A vector of claim 28, wherein the vector is a recombinant expression vector.

30. A vector of claim 28, wherein the vector is a viral vector.

31. A host cell comprising a polynucleotide encoding the polypeptide of claim 1 operably linked to a heterologous promoter.

32. A host cell comprising a vector of claim 28.

33. The host cell of claim 32, wherein the host cell is a stem cell.

34. The host cell of claim 32, wherein the host cell is a neuronal lineage cell.

35. An antibody or antigen binding fragment thereof that specifically binds the polypeptide of claim 1.

36. The antibody or antigen binding fragment thereof of claim 35 which is human or humanized.

37. The antibody or antigen binding fragment thereof of claim 35 which is an intrabody.

38. The antibody or antigen binding fragment thereof of claim 35 which specifically binds the N-terminal MICAL domain of said polypeptide.

39. The antibody or antigen binding fragment thereof of claim 35 which specifically binds plexin interacting region of said polypeptide.

40. A method for identifying an agent that affects axonal guidance regulatory activity, comprising contacting an isolated polypeptide of claim 1, or a cell recombinantly expressing a polypeptide of claim 1, with a candidate agent, and comparing the axonal guidance regulatory activity in the presence and absence of the agent, wherein a difference in activity is indicative of an agent that affects axonal guidance regulatory activity.

41. The method of claim 40, wherein the agent inhibits axonal guidance regulatory activity.

42. The method of claim 40, wherein the agent is a small molecule, an antisense polynucleotide, a MICAL-like polypeptide or fragment thereof, a mutant MICAL polypeptide, an anti-MICAL antibody, a double stranded RNA, or a peptidomimetic.

43. The method of claim 40, wherein the agent is a monooxygenase inhibitor.

44. The method of claim 43, wherein the anti-oxidant is a flavonoid.

45. The method of claim 44, wherein the flavonoid is a gallic acid derivative.

46. A method for screening for an agent that modulates an activity of a MICAL polypeptide, said method comprising (a) contacting the isolated polypeptide of claim 1 with a candidate agent and (b) comparing said activity of the polypeptide of claim 1 in the presence or absence of said candidate agent, wherein a difference in said activity indicates that the agent modulates the activity of the MICAL polypeptide.

47. The method of claim 46, wherein said activity is monooxygenase activity.

48. The method of claim 46, wherein said activity is plexin A-binding activity.

49. The method of claim 46, wherein the method is a cell-free assay.

50. A method for screening for an agent that modulates an activity of a MICAL polypeptide, said method comprising (a) contacting a cell expressing the polypeptide of claim 1 with a candidate agent and (b) comparing said activity of the polypeptide of claim 1 in the presence or absence of said candidate agent, wherein a difference in said activity indicates that the agent modulates the activity of the MICAL polypeptide.

51. The method of claim 50, wherein the activity is monooxygenase activity.

52. The method of claim 50, wherein the activity is plexin A-binding activity.

53. The method of claim 50, wherein the cell is a neuron.

54. The method of claim 50, wherein the cell is an immune cell.

55. The method of claim 50, wherein the cell has a transformed phenotype.

56. The method of claim 50, wherein the cell is a cardiac cell.

57. A method for screening for an agent that modulates an activity of a MICAL polypeptide, said method comprising (a) contacting a cell that recombinantly expresses the polypeptide of claim 1 with a candidate agent and (b) comparing a phenotypic or physiological trait of said cell in the presence or absence of said candidate agent, wherein a difference in said phenotypic or physiological trait indicates that the agent modulates the activity of the MICAL polypeptide.

58. The method of claim 57, wherein the phenotypic or physiological trait involves dynamics of the cytoskeleton.

59. The method of claim 57, wherein the phenotypic or physiological trait is axon guidance.

60. The method of claim 57, wherein the phenotypic or physiological trait is cell proliferation or invasiveness.

61. The method of claim 57, wherein the phenotypic or physiological trait is an immune response.

62. A method for screening for an agent that modulates the expression of a MICAL polypeptide, the method comprising (a) contacting a cell with a candidate agent; and (b) comparing the expression of the polypeptide of claim 1 in the presence or absence of the candidate agent, wherein a difference in the expression indicates that the agent modulates the expression of the MICAL polypeptide.

63. The method of claim 62, wherein the level of mRNA encoding MICAL is compared.

64. The method of claim 62, wherein the level of the MICAL polypeptide is compared.

65. A polynucleotide that specifically hybridizes to a polynucleotide of claim 15, wherein the polynucleotide is at least 15 nucleotides in length.

66. A polynucleotide of claim 65, wherein the polynucleotide inhibits expression of a polynucleotide that encodes a polypeptide of claim 1.

67. A polynucleotide of claim 65, wherein the polynucleotide is at least 90% identical to a complementary polynucleotide of a polynucleotide encoding a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

68. A polynucleotide of claim 65, wherein the polynucleotide specifically hybridizes to a polynucleotide encoding a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

69. A double-stranded RNA molecule comprising a first RNA strand that specifically hybridizes to an mRNA encoding a MICAL polypeptide and a second RNA strand that is the reverse complement of said first strand, wherein said double-stranded RNA molecule is at least 15 base pairs in length.

70. An isolated polypeptide or a functional peptide portion thereof, comprising a calponin homology domain, a LIM domain, a proline rich region, and a plexin interacting region, and having plexin-interacting activity.

71. An isolated polypeptide of claim 70, wherein the polypeptide comprises a calponin homology domain, followed by a first variable region, followed by a LIM domain, followed by a proline rich region, and followed by a plexin interacting region.

72. An isolated polypeptide of claim 70, wherein the polypeptide is at least 90% identical to an amino acid sequence as set forth in SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

73. An isolated polypeptide of claim 72, wherein the polypeptide has an amino acid sequence as set forth in SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

74. An isolated polypeptide of claim 70, wherein the polypeptide is a mammalian polypeptide.

75. An isolated polypeptide of claim 70, wherein the polypeptide is a human polypeptide.

76. An isolated polypeptide of claim 70, wherein the polypeptide is a *Drosophila* polypeptide.

77. An isolated polynucleotide encoding a polypeptide according to claim 70, or a functional peptide portion thereof.

78. An isolated polynucleotide of claim 77, wherein the polynucleotide encodes a mammalian MICAL-like polypeptide.

79. An isolated polynucleotide of claim 77, wherein the polynucleotide encodes a polypeptide that is at least 90% identical to an amino acid sequence as set forth in SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

80. An isolated polynucleotide of claim 77, wherein the polynucleotide encodes a polypeptide comprising a calponin homology domain, followed by a first non-conserved region, followed by a LIM domain, followed by a second non-conserved region, followed by a proline rich region, and followed by a plexin interacting region.

81. An isolated polynucleotide of claim 79, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of SEQ ID NO:14, SEQ ID NO:16, or SEQ ID NO:18.

82. A vector comprising a polynucleotide of claim 77.

83. A vector of claim 75, wherein the vector is a recombinant expression vector.

84. A recombinant host cell comprising the polynucleotide of claim 77 operably linked to a heterologous promoter.

85. An isolated polynucleotide of claim 77, wherein the polynucleotide encodes a human polypeptide.

86. An isolated polynucleotide of claim 77, wherein the polynucleotide encodes a human polypeptide.

87. A non-human transgenic animal having a genome comprising a transgene comprising a nucleotide sequence encoding a MICAL polypeptide operably linked to a heterologous promoter, wherein the non-human transgenic animal expresses the transgenic polynucleotide in the central nervous system, and wherein expression levels of the transgenic polynucleotide are sufficient to effect an axonal guidance phenotype of the non-human organism.

88. The non-human transgenic animal of claim 87, wherein the non-human transgenic animal is a mouse.

89. The non-human transgenic animal of claim 87, wherein the MICAL polypeptide is ectopically expressed.

90. The non-human transgenic animal of claim 87, wherein the MICAL polypeptide is expressed at a greater level in one or more cells of the non-human transgenic animal than the MICAL polypeptide is expressed in comparable cells of a comparable non-human transgenic animal.

91. A non-human transgenic animal having a genome comprising a recombinantly inactivated nucleotide sequence encoding a MICAL polypeptide that has been recombinantly inactivated, wherein the non-human transgenic animal has an altered axon guidance phenotype.

92. The non-human transgenic animal of claim 91, wherein the non-human transgenic animal is a mouse.

93. The non-human transgenic animal of claim 91, wherein the non-human transgenic animal is heterozygous for the nucleotide sequence that has been inactivated.

94. The non-human transgenic animal of claim 91, wherein the non-human transgenic animal is homozygous for the nucleotide sequence that has been recombinantly inactivated.

95. A method for inhibiting axonal guidance regulatory activity, comprising contacting a cell that expresses a polypeptide of claim 1 with a monooxygenase inhibitor, thereby inhibiting axonal guidance regulatory activity.

96. The method of claim 95, wherein the monooxygenase inhibitor is a flavonoid.

97. The method of claim 96, wherein the flavonoid is a gallic acid derivative.

98. The method of claim 97, wherein the gallic acid derivative is (-)-epigallocatechin gallate (EGCG), (-)-epicatachin (EC), (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3"-*O*-methyl-EGCG, 3"-*O*-methyl-ECG, 3"-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, or *n*-cetyl gallate.

99. The method of claim 95, wherein the contacting is performed *in vitro*.

100. The method of claim 95, wherein the contacting is performed *in vivo*.

101. A method for affecting a semaphorin-mediated process, comprising contacting a cell that expresses a polypeptide of claim 1 with an effective amount of a monooxygenase inhibitor, thereby affecting the semaphorin-mediated process.

102. The method of claim 101, wherein the agent inhibits semaphorin 1a-PlexA-mediated repulsive axon guidance.

103. The method of claim 102, wherein the cell is a neuron.

104. The method of claim 102, wherein the cell is an immune cell, a cancer cell, or a cardiac cell.

105. The method of claim 101, wherein the monooxygenase inhibitor is a flavonoid.

106. The method of claim 105, wherein the flavonoid is a gallic acid derivative.

107. The method of claim 106, wherein the gallic acid derivative is (-)-epigallocatechin gallate (EGCG), (-)-epicatachin (EC), (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3”-*O*-methyl-EGCG, 3”-*O*-methyl-ECG, 3”-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, or *n*-cetyl gallate.

108. A method for treating a neurological condition in a subject in need thereof, the method comprising contacting in the subject, a cell of the central nervous system (CNS) or peripheral nervous system (PNS) having a disrupted axonal connection or a cell that affects axonal growth of the CNS cell or the PNS cell, with an amount of an agent that modulates the activity or expression of a MICAL polypeptide, the amount being effective to modulate axonal guidance or axon out-growth regulatory activity.

109. The method of claim 108, wherein the cell is contacted with the agent for a length of time effective for allowing axon regrowth.

110. The method of claim 108, wherein the agent is applied for a an length of time sufficient to promote neurorestoration.

111. The method of claim 108, wherein the agent is applied chronically.

112. The method of claim 108, wherein the neurological condition is a spinal cord injury.

113. The method of claim 108, wherein the neurological condition is traumatic brain injury.

114. The method of claim 108, wherein the neurological condition is neuropathic pain.

115. The method of claim 108, wherein the neurological condition is Parkinson's Disease.

116. The method of claim 108, wherein the neurological condition is Amyotrophic Lateral Sclerosis.

117. The method of claim 108, wherein the neurological condition is ischemic injury.

118. The method of claim 108, wherein the neurological condition is Alzheimer's Disease.

119. The method of claim 108, wherein the neurological condition is Multiple Sclerosis.

120. The method of claim 108, wherein the neurological condition is a neuropathy resulting from a stroke.

121. The method of claim 108, wherein the agent is an anti-oxidant.

122. The method of claim 121, wherein the anti-oxidant is a flavonoid.

123. The method of claim 122, wherein the flavonoid is a gallic acid derivative.

124. The method of claim 123, wherein the gallic acid derivative is (-)-epigallocatechin gallate (EGCG), (-)-epicatachin (EC), (-)-gallocatechin-3-*O*-gallate (GCG), (-)-epicatechin-3-*O*-gallate (ECG), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), theasinensin A, 3"-*O*-methyl-EGCG, 3"-*O*-methyl-ECG, 3"-*O*-methyl-GCG, (-)-epigallocatechin (EGC), (1)-gallocatechin (GC), gallic acid, catechin, *n*-octyl gallate, or *n*-cetyl gallate.

125. The method of claim 108, wherein the method further comprises contacting the cell with an agent that affects axon regeneration.

126. The method of claim 125, wherein the effect is promotion of axon regeneration.

127. The method of claim 126, wherein the agent is a neurotrophin, a mechanical bridge, or a stem cell.

128. The method of claim 127, wherein the mechanical bridge is a cell engineered to express a neurotrophic factor, a growth factor, an axon outgrowth promoting molecule, or an artificial polymer-based substrate.

129. The method of claim 127, wherein the mechanical bridge comprises fetal tissue.

130. The method of claim 127, wherein the mechanical bridge comprises a schwann cell or an olfactory ensheathing glia.

131. The method of claim 127, wherein the agent is a neurotrophic factor, a growth factor, an axon outgrowth promoting molecule, or an artificial polymer-based substrate.

132. The method of claim 131, wherein the agent is NGF, BDNF, NT-3, or NT-4/5.

133. The method of claim 131, wherein the agent is CNTF, GDNF, FGF, EGF, or PDGF.

134. The method of claim 131, wherein the agent is netrin, laminin, or collagen.

135. A method for treating a neurological disorder involving a failure of axon regrowth, comprising contacting a neuron having axons that fail to regrow, or surrounding tissue, with an agent that neutralizes oxidants for a length of time effective for allowing axon regrowth, thereby treating the neurological disorder.

136. The method of claim 135, wherein the agent is applied for a length of time sufficient to promote neurorestoration.

137. The method of claim 135, wherein the agent is applied for a length of time to diminish free radicals chronically after spinal cord injury.

138. The method of claim 135, wherein the agent is an anti-oxidant.

139. The method of claim 138, wherein the anti-oxidant is a flavonoid.

140. The method of claim 139, wherein the flavonoid is a gallic acid derivative.

141. The method of claim 140, wherein the gallic acid derivative is (-)-epigallocatechin gallate (EGCG) or (-)-epicatachin (EC).

142. A method for inducing regrowth and preventing inhibition of an injured process of a neuron, comprising altering the levels of reactive oxygen species in the milieu of the neuron for an length of time effective for allowing the injured process to reach a target.

143. The method of claim 142, wherein the levels of reactive oxygen species are altered for a length of time sufficient to promote neurorestoration.

144. The method of claim 142, further comprising identifying a site that includes the neuron suspected of having an injured process, before altering the levels of reactive oxygen species in the milieu of the neuron.

145. The method of claim 142, wherein the neuronal process is an axon.

146. The method of claim 142, wherein the neuronal process is a dendrite.

147. The method of claim 142, wherein levels of reactive oxygen species or other oxidation products are decreased.

148. The method of claim 142, wherein the method further comprises adding an agent that promote neuron process regrowth to the milieu of the neuron.

149. The method of claim 148, wherein the agent is a neurotrophin, a mechanical bridge, or a stem cell.

150. The method of claim 148, wherein levels of reactive oxygen species are lowered for at least 1 month.

151. The method of claim 148, wherein levels of reactive oxygen species are lowered for at least 3 months.

152. A method for limiting abnormal axon outgrowth, comprising contacting a neuron or the milieu of the neuron with an agent that affects oxidation state.

153. The method of claim 150, wherein the abnormal axon outgrowth is excessive axon outgrowth.

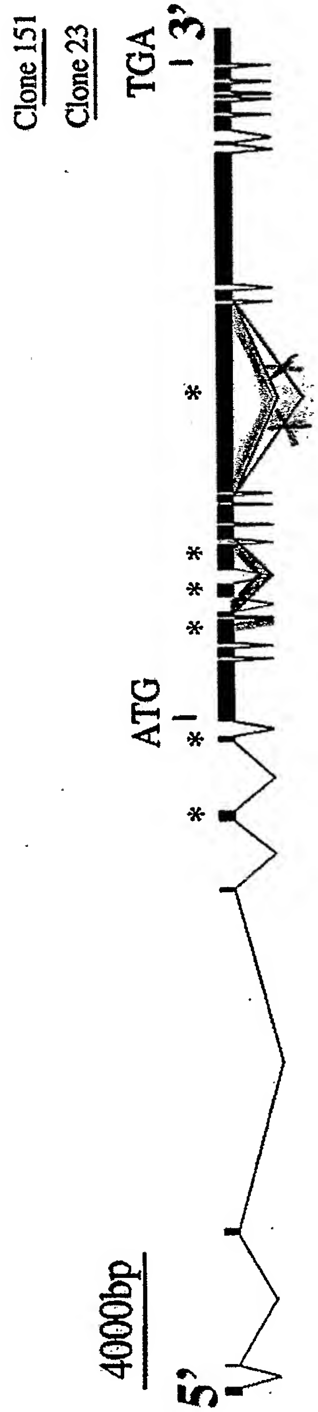


FIGURE 1A

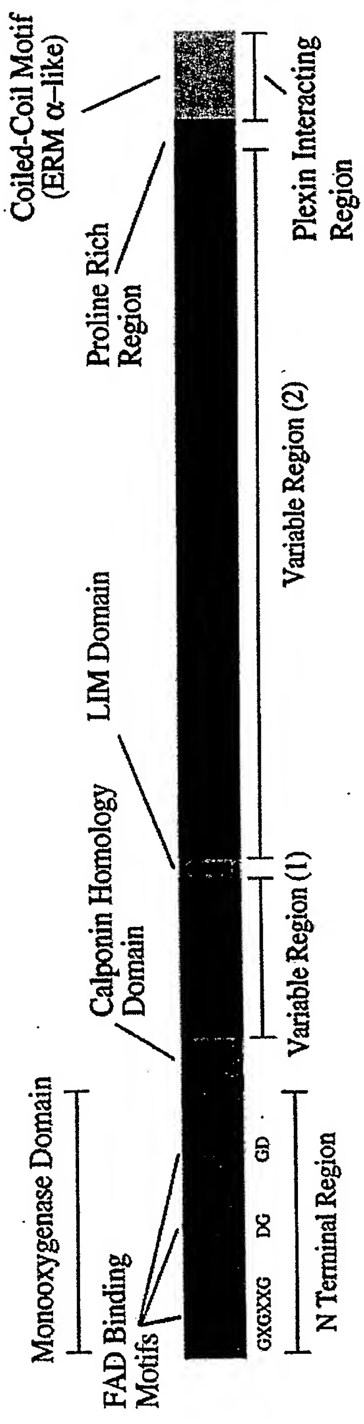


FIGURE 1B

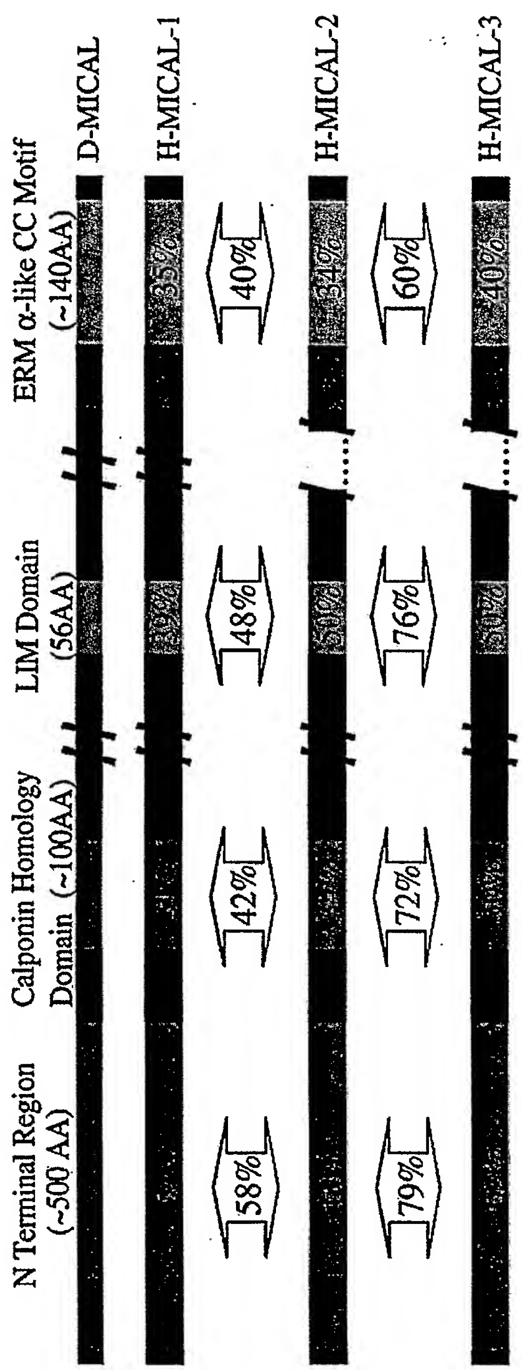


FIGURE 2A

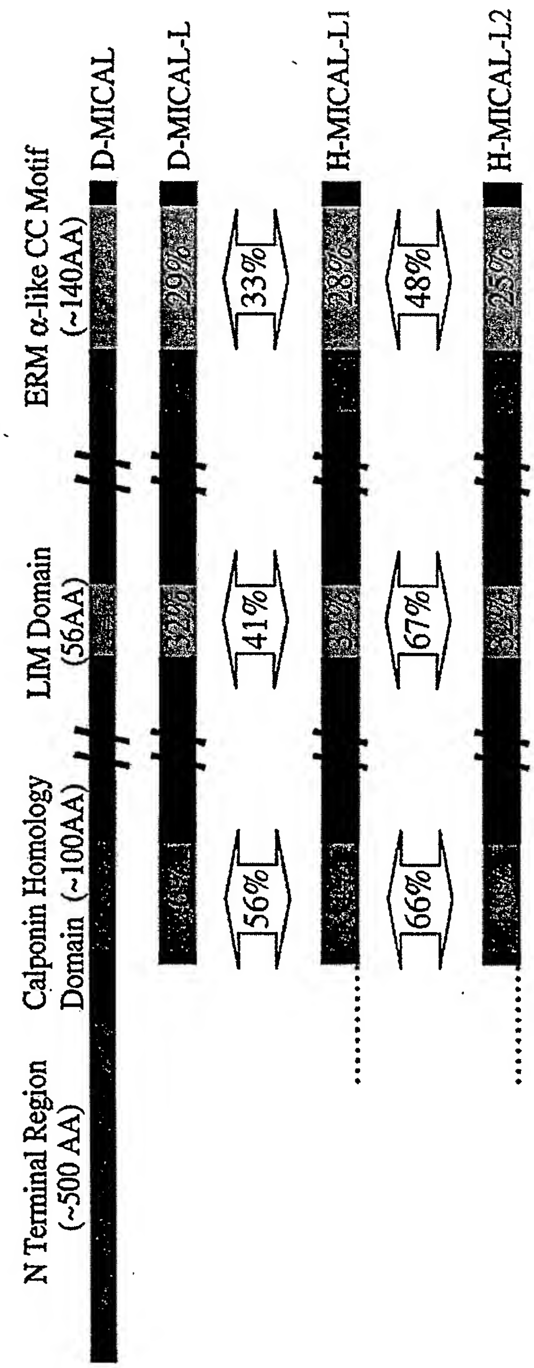


FIGURE 2B

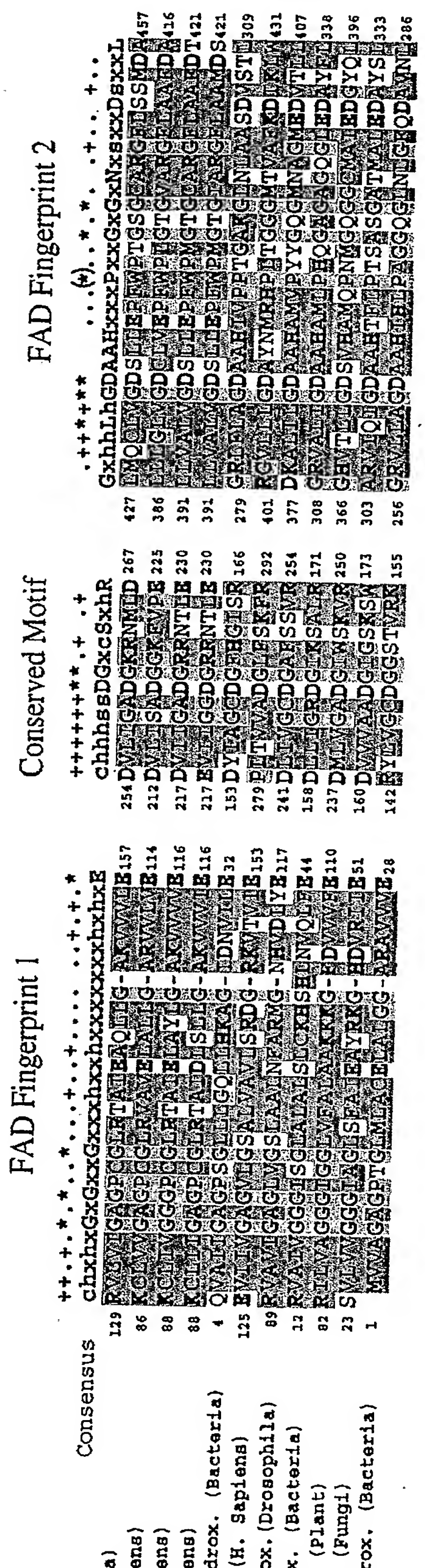


FIGURE 3A

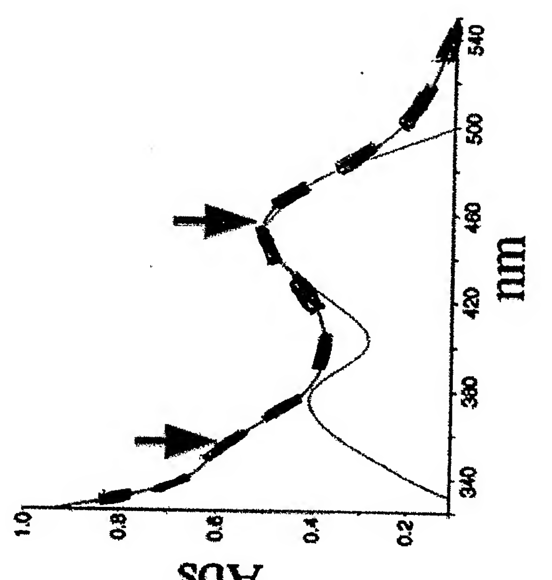


FIGURE 3B

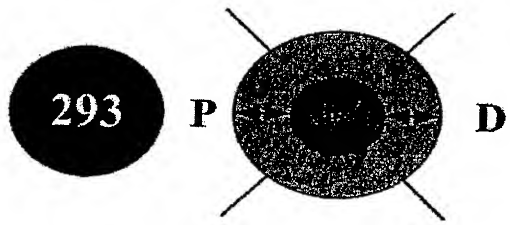


FIGURE 4A

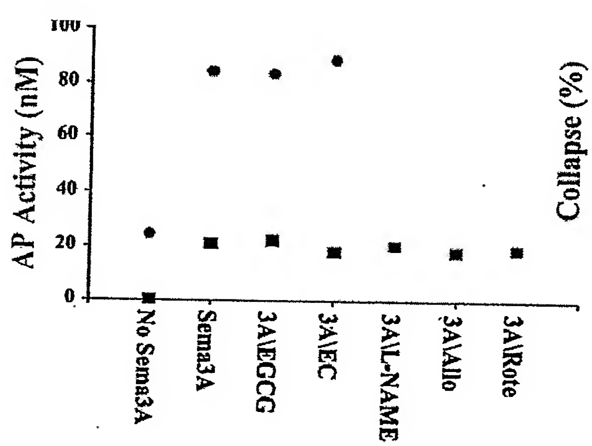


FIGURE 4B

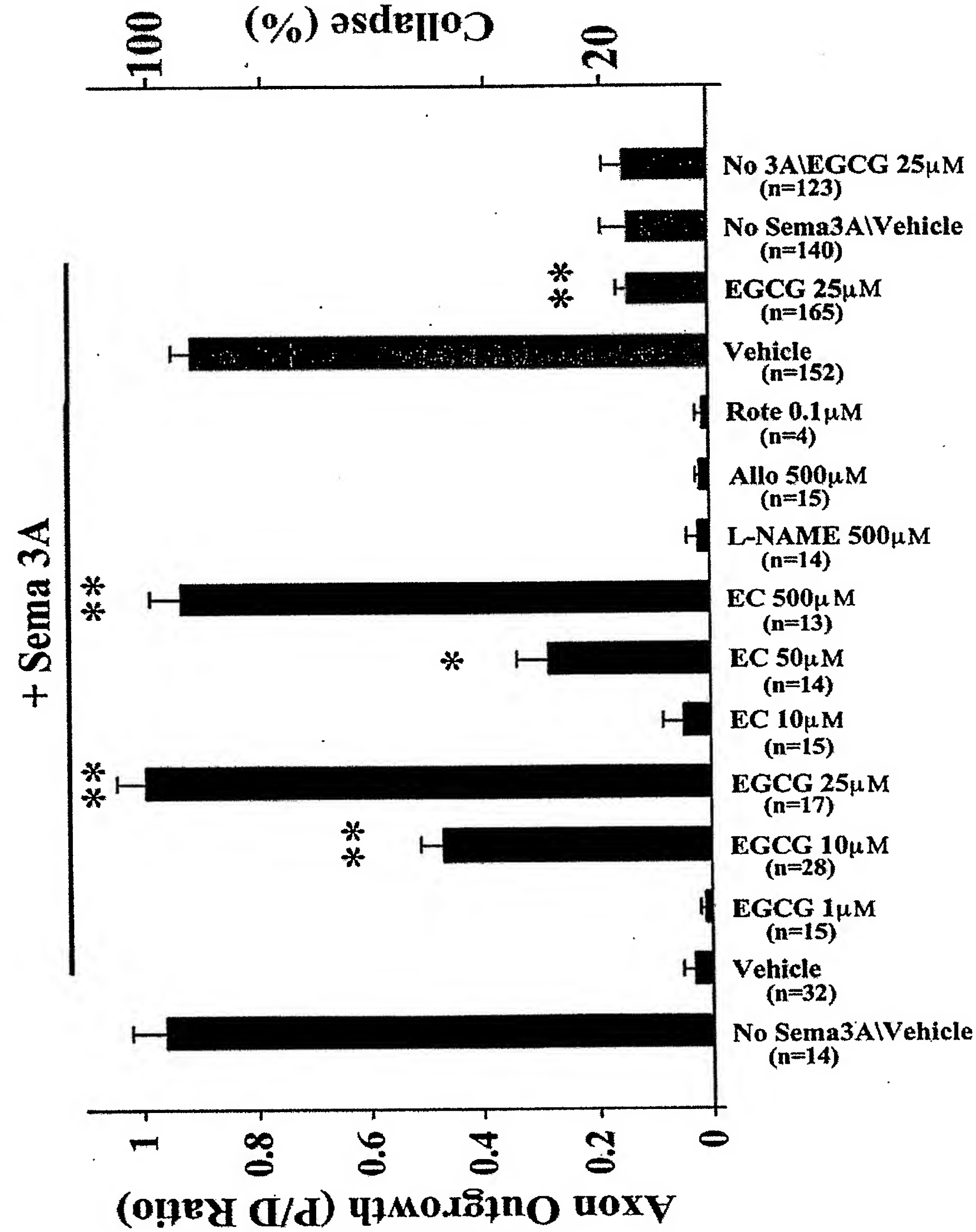

Collapse (%)
100
20
+ Sema 3A
No 3A\EGCG 25μM
(n=123)
No Sema3A\Vehicle
(n=140)
EGCG 25μM
(n=165)
Vehicle
(n=152)
Rote 0.1μM
(n=4)
Allo 500μM
(n=15)
L-NAME 500μM
(n=14)
EC 500μM
(n=13)
EC 50μM
(n=14)
EC 10μM
(n=15)
EGCG 25μM
(n=17)
EGCG 10μM
(n=28)
EGCG 1μM
(n=15)
Vehicle
(n=32)
No Sema3A\Vehicle
(n=14)
1
0.8
0.6
0.4
0.2
0
Axon Outgrowth (P/D Ratio)

FIGURE 4C

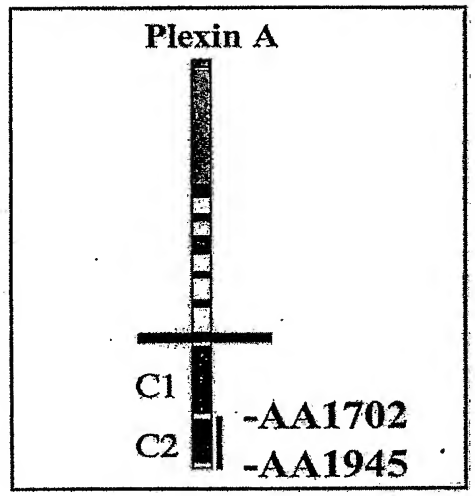


FIGURE 5A

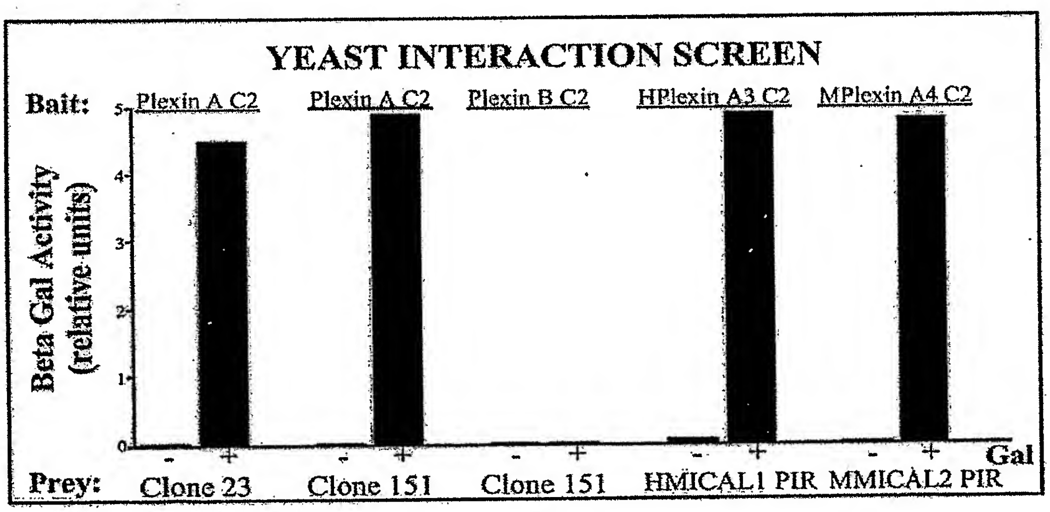


FIGURE 5B

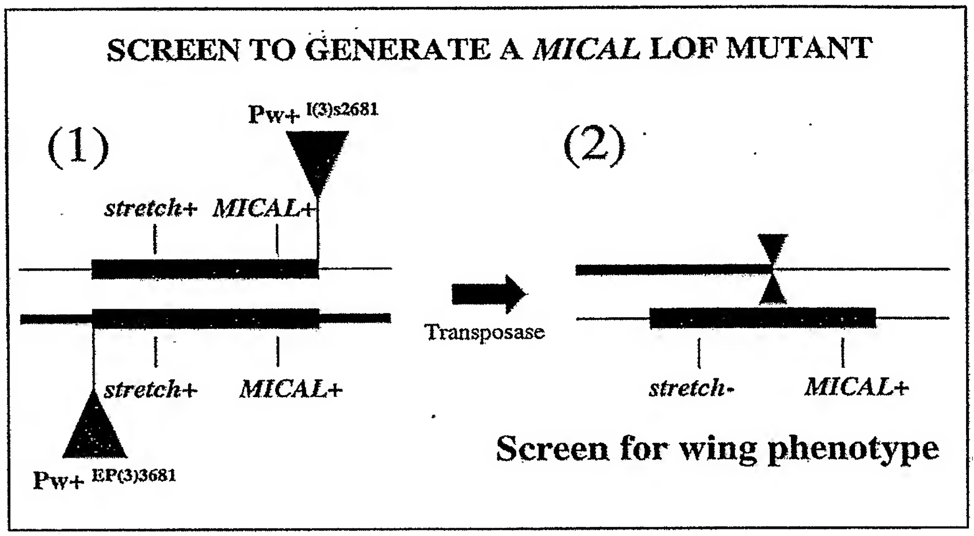


FIGURE 6A

GENETIC COMPLEMENTATION ANALYSES

Wing phenotype line/ Deficiency	swp2	swp4	swp7	swp11	swp38
l(3)10477 (*stretch*⁻)	–	–	–	–	–
Df(3R)by10	+	+	+	+	+
Df(3R)by62	–	–	–	–	–
Df(3R)segG16	+	+	+	+	+
Df(3R)mr73	–	+	–	–	–
Df(3R)mrO1	–	+	+	–	+

FIGURE 6B

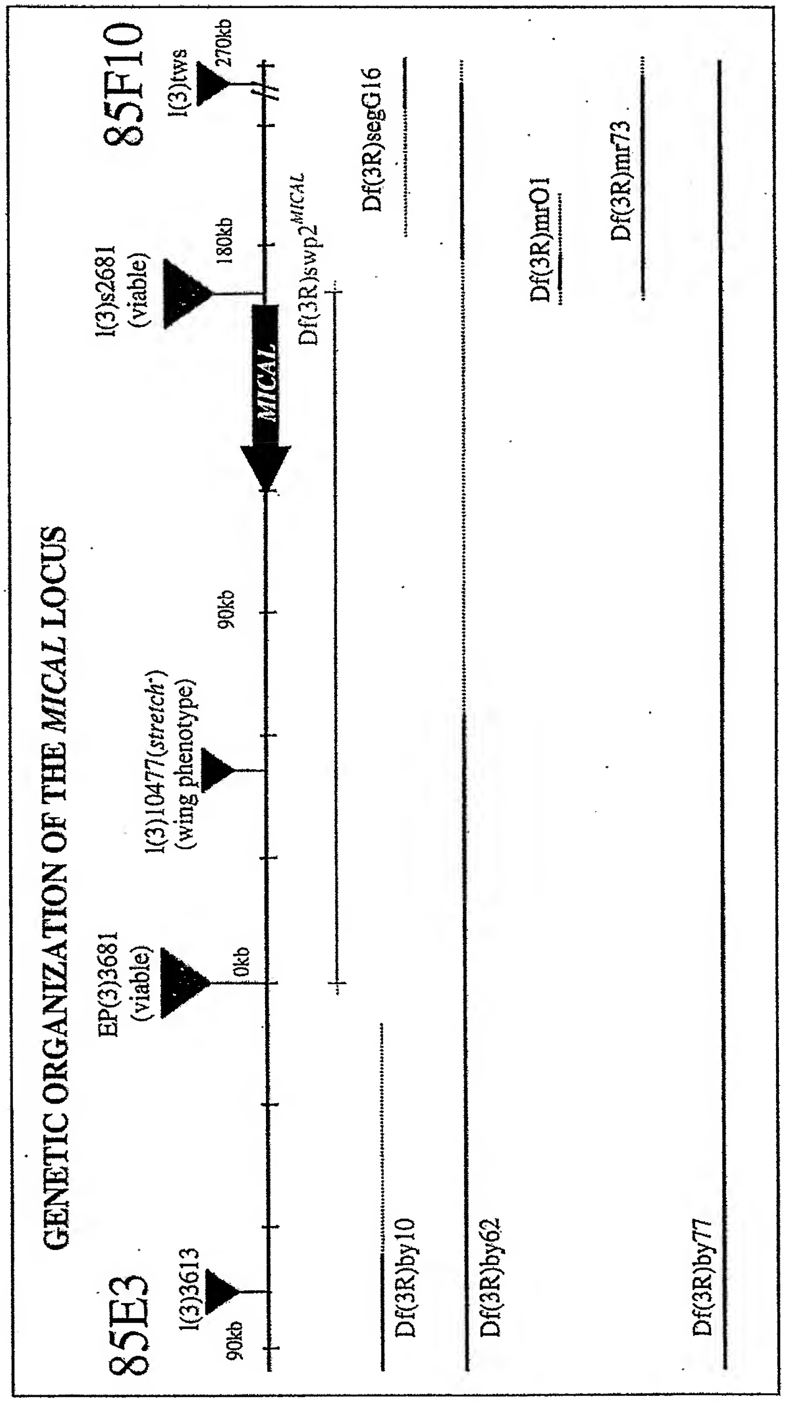

GENETIC ORGANIZATION OF THE MICAL LOCUS
85E3
85F10
l(3)3613
90kb
EP(3)3681
(viable)
0kb
l(3)10477(stretch)
(wing phenotype)
90kb
MICAL
l(3)s2681
(viable)
180kb
l(3)tws
270kb
Df(3R)swp2^MICAL
Df(3R)by10
Df(3R)segG16
Df(3R)by62
Df(3R)mrO1
Df(3R)mr73
Df(3R)by77

FIGURE 6C

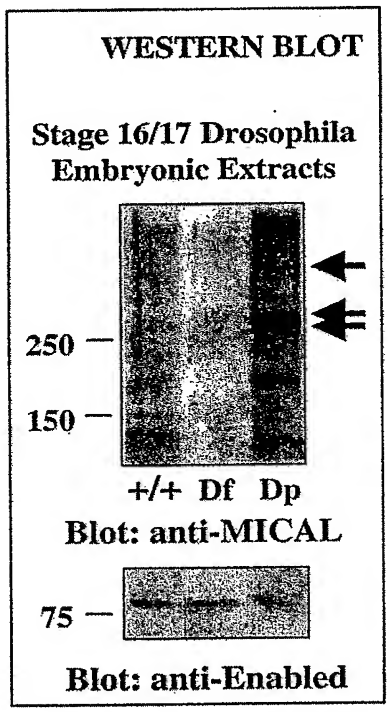

WESTERN BLOT
Stage 16/17 Drosophila
Embryonic Extracts
250
150
+/+ Df Dp
Blot: anti-MICAL
75
Blot: anti-Enabled

FIGURE 6D

MSRQHQRHHQQHHHLPPHQQPQQQMPQQQQQLTAQQQQQQQLLMAEHAAAAEAAELFDLL
CVATTMRQILALHRAMCEAVGLRPSPLNDFYPRLKAKVRSWKAQALWKKFDARAAHRVYG
KGAACTGTRVLVIGAGPCGLRTAIEAQLLGAKVVVLEKRDRITRNNVLHLWPFVITDLRNLGA
KKFYGKFCAGSIDHISIRQLQCMLLKVALLLGVEIHEGVSFDHAVEPSGDGGGWRAAVTPADH
PVSHYEFDVLIGADGKRNMLDFRRKEFRGKLAIAITANFINKKTEAEAKVEEISGVAFIFNQAFF
KELYGKTGIDLENIVYYKDETHYFVMTAKKHSLIDKGVIIEDMADPGELLAPANVDTQKLHD
YAREAAEFSTQYQMPNLEFAVNHYGKPDVAMFDFTSMFAAEMSCRVIVRKGARLMQCLVGD
SLLEPFWPTGSGCARGFLSSMDAAYAIKLWSNPQNSTLGVLAQRESIYRLLNQTTPDTLQRDIS
AYTVDPATRYPNLNRESVNSWQVKHLVDTDDPSILEQTFMDTHALQTPHLDTPGRRKRRSGD
LLPQGATLLRWISAQLHSYQFIPELKEASDVFRNGRVLCALINRYRPDLIDYAATKDMSPVECN
ELSFAVLERELHIDRVMSAKQSLDLTELESRIWLNYLDQICDLFR*GEIPHIKHPKMDFSDLRQKY*
*RINHTHAQPDFSKLLATKPKAKSPMQDAVDIPTTVQRRSVLEEERAKRQRRHEQ*LLNIGGGAAGA
AAGVAGSGTGTTTQGQNDTPRRSKKRRQVDKTANIEERQQRLQEIEENRQERMSKRRQQRCH
QTQNFYKSLQLLQAGKLLREGGEAGVAEDGTPFEDYSIFLYRQQAPVFNDRVKDLERKLLFPD
RERGDIPSALPRTADEQFSDRIKNMEQRMTGRGGLGGDKKPKDLMRAIGKIDSNDWNVREIE
KKIELSKKTEIHGPKGREKVPKWSKEQFQARQHKMSKPQRQDSREAEKFKDIDQTIRNLDKQL
KEGHNLDVGERGRNKVASIAGQFGKKDEANSDEKNAGSSNATTNTNNTVIPKSSSKVALAFK
KQAASEKCRFCKQTVYLMEKTTVEGLVLHRNCLKCHHCHTNLRLGGYAFDRDDPQGRFYCT
QHFRLPPKPLPQRTNKARKSAAAQPASPAVPPTAGSVPTAAATSEHMDTTPPRDQVDLLQTSR
ANASADAMSDDEANVIDEHEWSGRNFLPESNNDSQSELSSSDESDTESDSEMFEEADDSPFGA
QTLQLASDWIGKQYCEDSDDSDDFYDSSEGIADDGKDDTEGEEFKKARELRRQEVRLQPLPA
NLPTDTETEVQTESESTSPDEVELNSATEISTDSEFDNDEIIRQAPKIFIDDTHLRKPTKVQIKSTM
IGPNAASAGLHQKQLAAREKGGSYLQKYQPQPPLSQFKPLVQVDPTLLIGSQRAPLQNPRPGD
YLLNKTASTEGIASKKSLELKKRYLLGEPANGDKIQKSGSTSVLDSRIRSFQSNISECQKLLNPS
SDISAGMRTFLDRTKLGEGSQTTPGQTNELIRSATSNVINDLRVELRIQKTDSSHSTDNEKENVF
VNCKNELNKGMEYTDAVNATLLDQLARKSSPTTPTNKTVVEVIDLVTPEKPIDIIDLTALETPK
KQLVDGSAMDVDERLTPDSNKISELQQEVKEEPKPDVSRDVKECIPDILGHIKEGTGSKEPGGE
DQQSLLEQSDEEKRDSPEKDVAEHELYEPDSVQIQVPNIPWEKSKPEVMSTTGSSGSICSSSDSS
SIEDIQHYILESTTSPDTQTVGGKHNVPRLEVHDTSGALMQVDSLMIVNGKYIGDPEDVKFLD
MPANVIVPPAPALKTNELDMEDDQEAEAEPVTATPEPVECTVIEAERRVTAPPPLPEMGPPKLK
FDSKNENKIESLKNLPLIVESNVEHSQAVKPITLNLSNLARTPDTPTTPTAHDSDKTPTGEILSRG
SDSETEHTGTGQVLTETELSDWTADDCISENFVDLEFALNSNKGTIKRRKDRRRSGASKLPSGN
EVIHELARQAPVVQMDGILSAIDIDDIEFMDTGSEGSCAEAYPATNTALIQNRGYMEYIEAEPK
KTTRKAAPPSSYPGNLPPLMTKRDEKLGVDYIEQGAYIMHDDAKTPVNEVAPAMTQSLTDSIT
LNELDDDSMIISQTQPTTTEESEALTVVTSPLDTSSPRVLDQFASMLAAGKGDSTPSSSEQQPKT
STVTSSSTGPNSSTTGNVSKEPQEEDLQ*IQFEYVRALQQR*ISQISTQRRKSSKGEAPNLQLNSSAP
VIESAEDPAKPAEEPLVSMRPRTTSISGKVPEIPTLSSKLEEITKERTKQKDLIHDLVMDKLQSK
KQLNAEKRLHRSRQRSLLTSGYASGSSLSPTPKLAAACSPQDSNCSSQAHYHASTAEEAPKPP
AERPLQKSATSTYVSPYRTVQAPTRSADLYKPRPFSEHIDSNALAGYKLGKTASFNGGKLGDF
AKPIAPARVNRGGGVATADIANISASTENLRSEARARARLKSNTELGLSPEEKMQLIRSRLHYD
QNRSLKPKQLEEMPSGDLAARARKMSASKSVNDLAYMVGQQQQQQVEKDAVLQAKAADFT
SDPNLASGGQEKAGKTKSGRRPKDPERRKSLIQSLSSFFQKGSGSAASSSKEQGGAVAAVHSE
QSERPGTSSSGTPTISDAAGGGGGGGGVFSRFRISPKSKEKSKSCFDLRNFGFGDKDMLVCNAA
SPAGATSASQKNHSQEYLNTTNNSRYRKQTNTAKPKESFSSSPQLYIHKPHHLAAAHPSALD
DQT_PPPIPPLPL_**NYQRSDDESYANETREHKKQRAISKASRQAELKRLRIAQEIQREQEEIEV**
QLKDLEARGVLIEKALRGEAQNIENLDATKDNDEKLLKELLEIWRNITALKKRDEELTIR
QQELQLEYRHAQLKEELNLRLSCNKLDKSSADVAAEGAILNEMLEIVAKRAALRPTASQ
LDLTAAGSASTSAEATGIKLTGQPHDHEESII

FIGURE 7

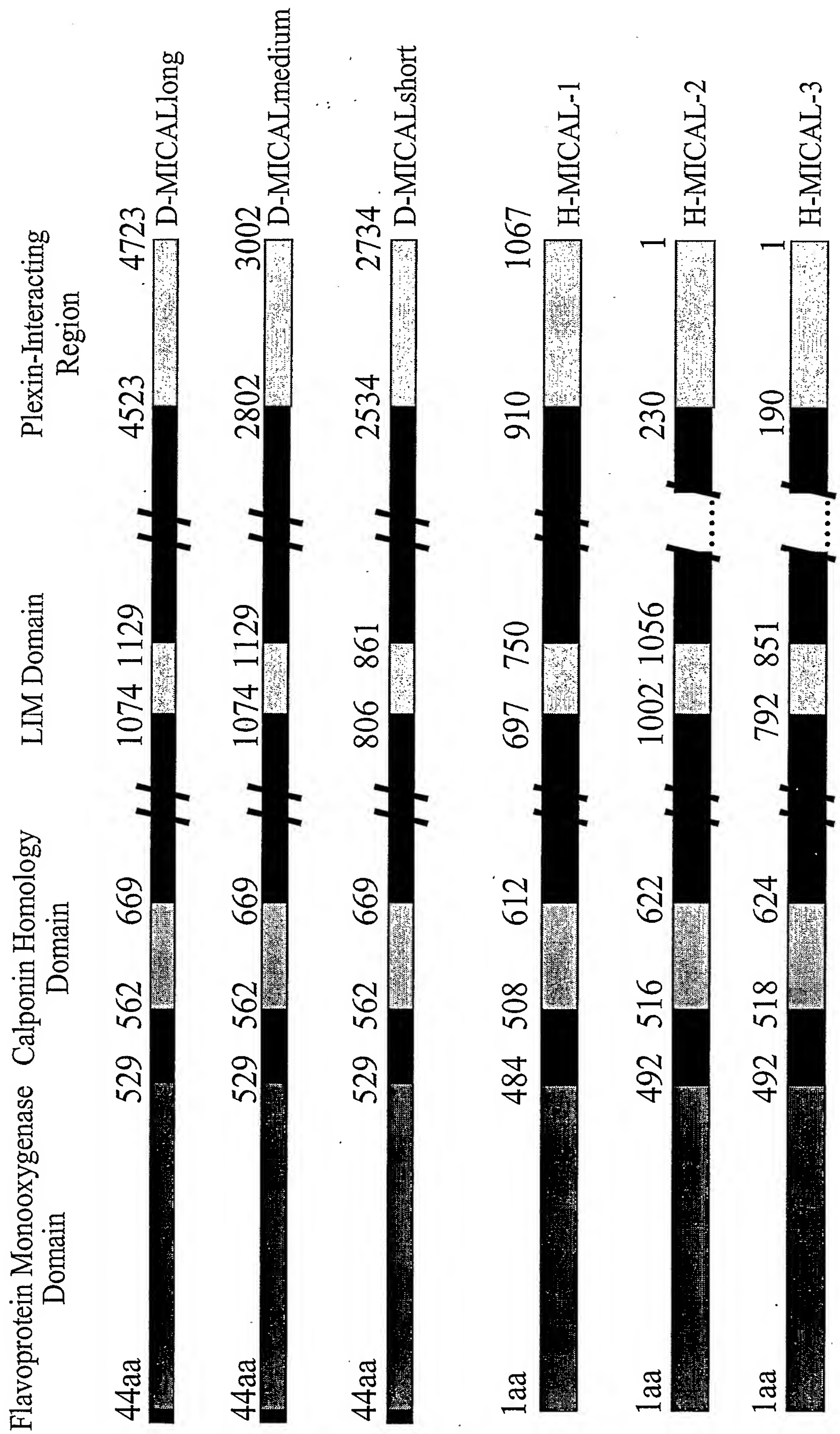

Flavoprotein Monooxygenase Domain
Calponin Homology Domain
LIM Domain
Plexin-Interacting Region
44aa 529 562 669 1074 1129 4523 4723 D-MICALlong
44aa 529 562 669 1074 1129 2802 3002 D-MICALmedium
44aa 529 562 669 806 861 2534 2734 D-MICALshort
1aa 484 508 612 697 750 910 1067 H-MICAL-1
1aa 492 516 622 1002 1056 230 1 H-MICAL-2
1aa 492 518 624 792 851 190 1 H-MICAL-3

FIGURE 8

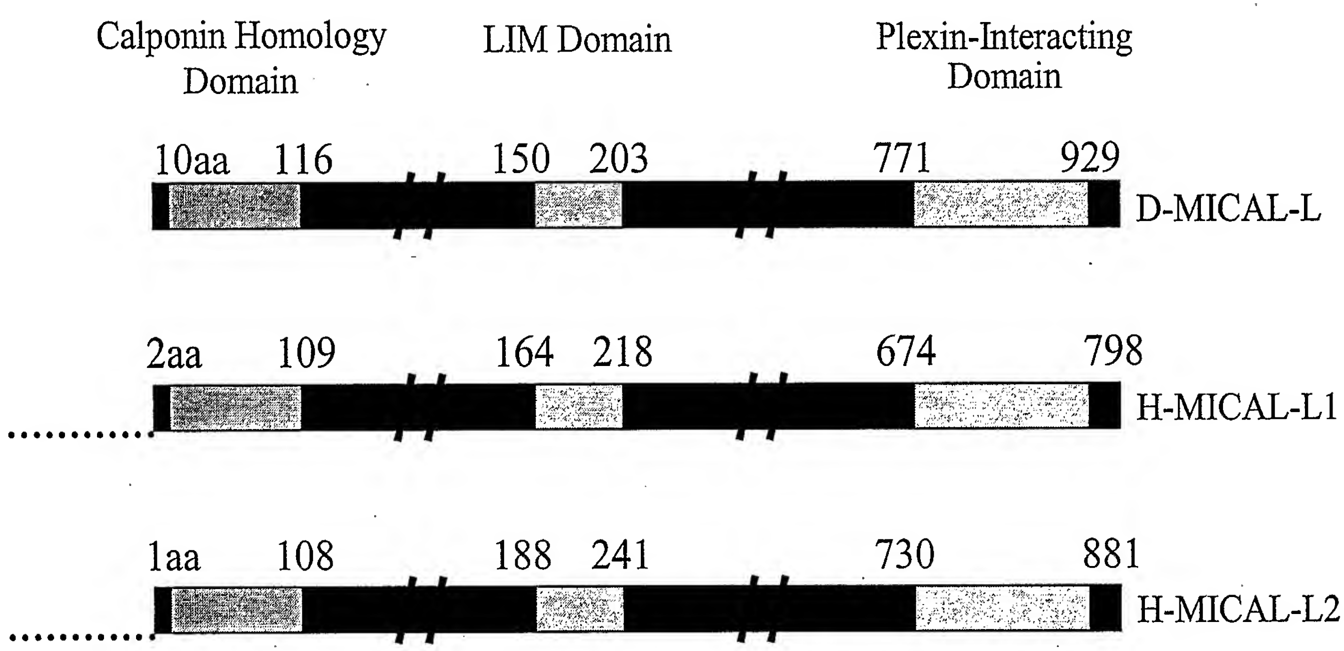

Calponin Homology Domain
LIM Domain
Plexin-Interacting Domain
10aa
116
150
203
771
929
D-MICAL-L
2aa
109
164
218
674
798
H-MICAL-L1
1aa
108
188
241
730
881
H-MICAL-L2

FIGURE 9

MASPASTNPAHDHFETFVQAQLCQDVLSSFQGLCRALGVESGGGLSQYHKIKAQ
LNYWSAKSLWAKLDKRASQPVYQQGQACTNTKCLVVGAGPCGLRAAVELALL
GARVVLVEKRIKFSRHNVLHLWPFTIHDLRALGAKKFYGRFCTGTLDHISIRQLQ
LLLLKVALLLGVEIHWGVKFTGLQPPPRKGSGWRAQLQPNPPAQLASYEFDVLIS
AAGGKFVPEGFTIREMRGKLAIGITANFVNGRTVEETQVPEISGVARIYNQKFFQS
LLKATGIDLENIVYYKDETHYFVMTAKKQCLLRLGVLRQDLSETDQLLGKANVV
PEALQRFARAAADFATHGKLGKLEFAQDARGRPDVAAFDFTSMMRAESSARVQ
EKHGARLLLGLVGDCLVEPFWPLGTGVARGFLAAFDAAWMVKRWAEGAGPLE
VLAERESLYQLLSQTSPENMHRNVAQYGLDPATRYPNLNLRAVTPNQVQDLYD
MMDKEHAQRKSDEPDSRKTTTGSAGTEELLHWCQEQTAGFPGVHVTDFSSSWA
DGLALCALVHHLQPGLLEPSELQGMGALEATTWALRVAEHELGITPVLSAQAV
MAGSDPLGLIAYLSHFHSAFKNTSHSSGLVSQPSGTPSAILFLGKLQRSLQRTRAK
VDEETPSTEEPPVSEPSMSPNTPELSEHQEAGAEELCELCGKHLYILERFCVDGHF
FHRSCFCCHTCEATLWPGGYGQHPGDGHFYCLQHLPQEDQKEADNNGSLESQE
LPTPGDSNMQPDPSSPPVTRVSPVPSPSQPARRLIRLSSLERLRLSSLNIIPDSGAEP
PPKPPRSCSDLARESLKSSFVGWGVPVQAPQVPEAIEKGDDEEEEEEEEEEEEEPL
PPLEPELEQTLLTLAKNPGAMTKYPTW**RRTLMRRAKEEEMKRFCKAQAIQRR
LNEIEATMRELEAEGTKLELALRKESSSPEQQKKLWLDQLLRLIQKKNSLV
TEEAELMITVQELDLEEKQRQLDHELRGYMNREETMKTEADLQSENQVLR
KLLEVVNQRDALIQFQEERRLREMPA**

FIGURE 10

HMICAL-1.txt

```
ccacctctccagccactcatctctgcccagctgctgccctccccaggaggcctcc
1/1                                     31/11                                   61/21                                   91/31
atg gct tca cct acc tcc acc aac cca gcg cat gcc cac ttt gag agc ttc ctg cag gcc cag ctg tgc cag gac gtg ctg agc agc ttc cag gag ctg tgt ggg gcc ctg ggg ctg gaa
M   A   S   P   T   S   T   N   P   A   H   A   H   F   E   S   F   L   Q   A   Q   L   C   Q   D   V   L   S   S   F   Q   E   L   C   G   A   L   G   L   E
121/41                                  151/51                                  181/61                                  211/71
ccc ggt ggg ggg ctg ccc cag tac cac aag atc aag gac cag ctc aac tac tgg agc gcc aag tca ctg tgg acc aag ctg gac aag cga gca ggc cag cct gtc tac cag cag ggc cgg
P   G   G   G   L   P   Q   Y   H   K   I   K   D   Q   L   N   Y   W   S   A   K   S   L   W   T   K   L   D   K   R   A   G   Q   P   V   Y   Q   Q   G   R
241/81                                  271/91                                  301/101                                 331/111
gcc tgc acc agc acc aag tgc ctg gtg gtg ggt gct gga cct tgc ggg ctg cgg gtc gct gtg gag ctg gcg ctg ctg ggg gcc cga gtg gtg ctg gtg gaa aag cgc acc aag ttc tct
A   C   T   S   T   K   C   L   V   V   G   A   G   P   C   G   L   R   V   A   V   E   L   A   L   L   G   A   R   V   V   L   V   E   K   R   T   K   F   S
361/121                                 391/131                                 421/141                                 451/151
cgc cac aac gtg ctc cac ctc tgg ccc ttc acc atc cac gac ctg cgg gca ctc ggt gct aag aag ttc tac ggg cgc ttc tgc acc ggc acc ctg gac cac atc agc atc agg cag ctc
R   H   N   V   L   H   L   W   P   F   T   I   H   D   L   R   A   L   G   A   K   K   F   Y   G   R   F   C   T   G   T   L   D   H   I   S   I   R   Q   L
481/161                                 511/171                                 541/181                                 571/191
cag ctg ctt ctg ctg aag gta gca ttg ctg ctg ggg gtg gaa att cac tgg ggt gtc act ttc act ggc ctc cag ccc cct cct agg aag ggg agt ggc tgg cgt gcc cag ctc caa ccc
Q   L   L   L   L   K   V   A   L   L   L   G   V   E   I   H   W   G   V   T   F   T   G   L   Q   P   P   P   R   K   G   S   G   W   R   A   Q   L   Q   P
601/201                                 631/211                                 661/221                                 691/231
aac ccc cct gcc cag ctg gcc aac tat gaa ttt gac gtc ctt atc tcg gct gca gga ggt aaa ttc gtc cct gaa ggc ttc aaa gtt cga gaa atg cga ggc aaa ctg gcc att ggc atc
N   P   P   A   Q   L   A   N   Y   E   F   D   V   L   I   S   A   A   G   G   K   F   V   P   E   G   F   K   V   R   E   M   R   G   K . L   A   I   G   I
721/241                                 751/251                                 781/261                                 811/271
aca gcc aac ttt gtg aat gga cgc acc gtg gag gag aca cag gtg ccg gag atc agt ggt gta gcc agg atc tac aac cag agc ttc ttc cag agc ctt ctc aaa gcc aca ggc att gat
T   A   N   F   V   N   G   R   T   V   E   E   T   Q   V   P   E   I   S   G   V   A   R   I   Y   N   Q   S   F   F   Q   S   L   L   K   A   T   G   I   D
841/281                                 871/291                                 901/301                                 931/311
ctg gag aac att gtg tac tac aag gac gac acc cac tac ttt gtg atg aca gcc aag aag cag tgc ctg ctg cgg ctg ggg gtg ctg cgc cag gac tgg cca gac acc aat cgg ctg ctg
L   E   N   I   V   Y   Y   K   D   D   T   H   Y   F   V   M   T   A   K   K   Q   C   L   L   R   L   G   V   L   R   Q   D   W   P   D   T   N   R   L   L
961/321                                 991/331                                 1021/341                                1051/351
ggc agt gcc aat gtg gtg ccc gag gct ctg cag cgc ttt acc cgg gca gct gct gac ttt gcc acc cat ggc aag ctc ggg aaa cta gag ttt gcc cag gat gcc cat ggg cag cct gat
G   S   A   N   V   V   P   E   A   L   Q   R   F   T   R   A   A   A   D   F   A   T   H   G   K   L   G   K   L   E   F   A   Q   D   A   H   G   Q   P   D
1081/361                                1111/371                                1141/381                                1171/391
gtc tct gcc ttt gac ttc acg agc atg atg cgg gca gag agt tct gct cgt gtg caa gag aag cat ggc gcc cgc ctg ctg ctg gga ctg gtg ggg gac tgc ctg gtg gag ccc ttc tgg
V   S   A   F   D   F   T   S   M   M   R   A   E   S   S   A   R   V   Q   E   K   H   G   A   R   L   L   L   G   L   V   G   D   C   L   V   E   P   F   W
1201/401                                1231/411                                1261/421                                1291/431
ccc ctg ggc act gga gtg gca cgg ggc ttc ctg gca gcc ttt gat gca gcc tgg atg gtg aag cgg tgg gca gag ggc gct gag tcc cta gag gtg ttg gct gag cgt gag agc ctg tac
P   L   G   T   G   V   A   R   G   F   L   A   A   F   D   A   A   W   M   V   K   R   W   A   E   G   A   E   S   L   E   V   L   A   E   R   E   S   L   Y
1321/441                                1351/451                                1381/461                                1411/471
cag ctt ctg tca cag aca tcc cca gaa aac atg cat cgc aat gtg gcc cag tat ggg ctg gac cca gcc acc cgc tac ccc aac ctg aac ctc cgg gca gtg acc ccc aat cag gta cga
Q   L   L   S   Q   T   S   P   E   N   M   H   R   N   V   A   Q   Y   G   L   D   P   A   T   R   Y   P   N   L   N   L   R   A   V   T   P   N   Q   V   R
1441/481                                1471/491                                1501/501                                1531/511
gac ctg tat gat gtg cta gcc aag gag cct gtg cag agg aac aac gac aag aca gat aca ggg atg cca gcc acc ggg tcg gca ggc acc cag gag gag ctg cta cgc tgg tgc cag gag
D   L   Y   D   V   L   A   K   E   P   V   Q   R   N   N   D   K   T   D   T   G   M   P   A   T   G   S   A   G   T   Q   E   E   L   L   R   W   C   Q   E
1561/521                                1591/531                                1621/541                                1651/551
cag aca gct ggg tac ccg gga gtc cac gtc tcc gat ttg tct tcc tcc tgg gct gat ggg cta gct ctg tgt gcc ctg gtg tac cgg ctg cag cct ggc ctg ctg gaa ccc tca gag ctg
Q   T   A   G   Y   P   G   V   H   V   S   D   L   S   S   S   W   A   D   G   L   A   L   C   A   L   V   Y   R   L   Q   P   G   L   L   E   P   S   E   L
1681/561                                1711/571                                1741/581                                1771/591
cag ggg ctg gga gct ctg gaa gca act gct tgg gca cta aag gtg gca gag aat gag ctg ggc atc aca ccg gtg gtg tct gca cag gcc gtg gta gca ggg agt gac cca ctg ggc ctc
Q   G   L   G   A   L   E   A   T   A   W   A   L   K   V   A   E   N   E   L   G   I   T   P   V   V   S   A   Q   A   V   V   A   G   S   D   P   L   G   L
1801/601                                1831/611                                1861/621                                1891/631
att gcc tac ctc agc cac ttc cac agt gcc ttc aag agc atg gcc cac agc cca ggc cct gtc agc cag gcc tcc cca ggg acc tcc agt gct gta tta ttc ctt agt aaa ctt cag agg
I   A   Y   L   S   H   F   H   S   A   F   K   S   M   A   H   S   P   G   P   V   S   Q   A   S   P   G   T   S   S   A   V   L   F   L   S   K   L   Q   R
1921/641                                1951/651                                1981/661                                2011/671
acc ctg cag cga tcc cgg gcc aag gaa aat gca gag gat gct ggt ggc aag aag ctg cgc ttg gag atg gag gcc gag acc cca agt act gag gtg cca cct gac cca gag cct ggt gta
T   L   Q   R   S   R   A   K   E   N   A   E   D   A   G   G   K   K   L   R   L   E   M   E   A   E   T   P   S   T   E   V   P   P   D   P   E   P   G   V
2041/681                                2071/691                                2101/701                                2131/711
ccc ctg aca ccc cca tcc caa cac cag gag gcc ggt gct ggg gac ctg tgt gca ctt tgt ggg gaa cac ctc tat gtc ctg gaa cgc ctc tgt gtc aac ggc cat ttc ttc cac cgg agc
P   L   T   P   P   S   Q   H   Q   E   A   G   A   G   D   L   C   A   L   C   G   E   H   L   Y   V   L   E   R   L   C   V   N   G   H   F   F   H   R   S
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 11-1

HMICAL-1.txt

```
tgc ttc cgc tgc cat acc tgt gag gcc aca ctg tgg cca ggt ggc tac gag cag cac cca gga gat gga cat ttc tac tgc ctc cag cac ctg ccc cag aca gac cac aaa gcg gaa ggc
C   F   R   C   H   T   C   E   A   T   L   W   P   G   G   Y   E   Q   H   P   G   D   G   H   F   Y   C   L   Q   H   L   P   Q   T   D   H   K   A   E   G
2281/761                                2311/771                                2341/781                                2371/791
agc gat aga ggc cct gag agt ccg gag ctc ccc aca cca agt gag aat agc atg cca cca ggc ctc tca act ccc aca gcc tcg cag gag ggg gcc ggt cct gtt cca gat ccc agc cag
S   D   R   G   P   E   S   P   E   L   P   T   P   S   E   N   S   M   P   P   G   L   S   T   P   T   A   S   Q   E   G   A   G   P   V   P   D   P   S   Q
2401/801                                2431/811                                2461/821                                2491/831
ccc acc cgt cgg cag atc cgc ctc tcc agc ccg gag cgc cag cgg ttg tcc tcc ctt aac ctt acc cct gac ccg gaa atg gag cct cca ccc aag cct ccc cgc agc tgc tcc gcc ttg
P   T   R   R   Q   I   R   L   S   S   P   E   R   Q   R   L   S   S   L   N   L   T   P   D   P   E   M   E   P   P   P   K   P   P   R   S   C   S   A   L
2521/841                                2551/851                                2581/861                                2611/871
gcc cgc cac gcc ctg gag agc agc ttt gtg ggc tgg ggc ctg cca gtc cag agc cct caa gct ctt gtg gcc atg gag aag gag gaa aaa gag agt ccc ttc tcc agt gaa gag gaa gaa
A   R   H   A   L   E   S   S   F   V   G   W   G   L   P   V   Q   S   P   Q   A   L   V   A   M   E   K   E   E   K   E   S   P   F   S   S   E   E   E   E
2641/881                                2671/891                                2701/901                                2731/911
gaa gat gtg cct ttg gac tca gat gtg gaa cag gcc ctg cag acc ttt gcc aag acc tca ggc acc atg aat aac tac cca aca tgg cgt cgg act ctg ctg cgc cgt gcg aag gag gag
E   D   V   P   L   D   S   D   V   E   Q   A   L   Q   T   F   A   K   T   S   G   T   M   N   N   Y   P   T   W   R   R   T   L   L   R   R   A   K   E   E
2761/921                                2791/931                                2821/941                                2851/951
gag atg aag agg ttc tgc aag gcc cag acc atc caa cgg cga cta aat gag att gag gct gcc ttg agg gag cta gag gcc gag ggc gtg aag ctg gag ctg gcc ttg agg cgc cag agc
E   M   K   R   F   C   K   A   Q   T   I   Q   R   R   L   N   E   I   E   A   A   L   R   E   L   E   A   E   G   V   K   L   E   L   A   L   R   R   Q   S
2881/961                                2911/971                                2941/981                                2971/991
agt tcc cca gaa cag caa aag aaa cta tgg gta gga cag ctg cta cag ctc gtt gac aag aac aac agc ctg gtg gct gag gag gcc gag ctc atg atc acg gtg cag gaa ttg aat ctg
S   S   P   E   Q   Q   K   K   L   W   V   G   Q   L   L   Q   L   V   D   K   N   N   S   L   V   A   E   E   A   E   L   M   I   T   V   Q   E   L   N   L
3001/1001                               3031/1011                               3061/1021                               3091/1031
gag gag aaa cag tgg cag ctg gac cag gag cta cga ggc tac atg aac cgg gaa gaa aac cta aag aca gct gct gat cgg cag gct gag gac cag gtc ctg agg aag ctg gtg gat ttg
E   E   K   Q   W   Q   L   D   Q   E   L   R   G   Y   M   N   R   E   E   N   L   K   T   A   A   D   R   Q   A   E   D   Q   V   L   R   K   L   V   D   L
3121/1041                               3151/1051                               3181/1061
gtc aac cag aga gat gcc ctc atc cgc ttc cag gag gag cgc agg ctc agc gag ctg gcc ttg ggg aca ggg gcc cag ggc tag acg agg gtg ggc cgt ctg ctt tcg ttc cca caa aga
V   N   Q   R   D   A   L   I   R   F   Q   E   E   R   R   L   S   E   L   A   L   G   T   G   A   Q   G   *
aag cac ctc acc cca gca cag tgc cac ccc tgt tca tct ggg ctg cct ggc aga gag cct tgc tgt tta caa tta aaa tgt ttc tgc cac
```

FIGURE 11-2

```
HMICAL-2.txt
ccgggccgcctcgctcgctcccagctctgtcagtggcccgcggggcccgatcgctgcgcccgcggccagggccgaggcaggcctgacccggggccgggcagcccgcgcgactttcggaacatggcaacccgtgtgtgtctcatcccagaaagagaagact
ttaaccactgtgatgcctgagaatccagtgtgacgtttctccagatacttcatgctgttcacctgtgtcctcgccgcaccactgccgcacacgactcctgaacc
1/1                                     31/11                                   61/21                                   91/31
atg ggg gaa aac gag gat gag aag cag gcc cag gcg ggg cag gtt ttt gag aac ttt gtc cag gca tcc acg tgc aaa ggt acc ctc cag gcc ttc aac att ctc aca cga cac ctg gac
M   G   E   N   E   D   E   K   Q   A   Q   A   G   Q   V   F   E   N   F   V   Q   A   S   T   C   K   G   T   L   Q   A   F   N   I   L   T   R   H   L   D
121/41                                  151/51                                  181/61                                  211/71
cta gac cct ctg gac cac aga aac ttt tat tcc aag ctc aag tcc aag gtg acc acc tgg aaa gcc aaa gcc ctg tgg tac aaa ttg gat aag cgt ggt tcc cac aaa gag tat aag cga
L   D   P   L   D   H   R   N   F   Y   S   K   L   K   S   K   V   T   T   W   K   A   K   A   L   W   Y   K   L   D   K   R   G   S   H   K   E   Y   K   R
241/81                                  271/91                                  301/101                                 331/111
ggg aag tcg tgc acg aac acc aag tgt ctc ata gtt ggg gga gga ccc tgt ggc ttg cgc act gcc att gaa ctt gcc tac ctg gga gcc aaa gtg gtc gtg gtg gag aag agg gac tcc
G   K   S   C   T   N   T   K   C   L   I   V   G   G   G   P   C   G   L   R   T   A   I   E   L   A   Y   L   G   A   K   V   V   V   V   E   K   R   D   S
361/121                                 391/131                                 421/141                                 451/151
ttc tcc cgg aac aac gtg cta cac ctc tgg cct ttc acc atc cat gac ctt cgg ggc ctg gga gcc aag aag ttc tat ggg aag ttc tgt gct ggc tcc atc gac cat atc agt att cgc
F   S   R   N   N   V   L   H   L   W   P   F   T   I   H   D   L   R   G   L   G   A   K   K   F   Y   G   K   F   C   A   G   S   I   D   H   I   S   I   R
481/161                                 511/171                                 541/181                                 571/191
caa cta cag ctc atc cta ttc aag gtg gcc ctg atg ctg gga gtt gaa atc cat gtg aat gtg gag ttc gtg aag gtt cta gag cct cct gaa gat caa gaa aat caa aaa att ggc tgg
Q   L   Q   L   I   L   F   K   V   A   L   M   L   G   V   E   I   H   V   N   V   E   F   V   K   V   L   E   P   P   E   D   Q   E   N   Q   K   I   G   W
601/201                                 631/211                                 661/221                                 691/231
cgg gca gaa ttt ctc cct aca gac cat tct ctg tcg gag ttt gag ttt gac gtc atc att ggt gcc gat ggc cgc agg aac acc ctg gaa ggg ttc aga aga aaa gaa ttc cgt ggg aag
R   A   E   F   L   P   T   D   H   S   L   S   E   F   E   F   D   V   I   I   G   A   D   G   R   R   N   T   L   E   G   F   R   R   K   E   F   R   G   K
721/241                                 751/251                                 781/261                                 811/271
ctg gcg att gcc atc acc gcc aac ttc ata aac aga aac agc aca gcg gaa gcc aag gtg gaa gag att agt ggt gtg gct ttc atc ttc aat cag aaa ttt ttt cag gac ctt aaa gaa
L   A   I   A   I   T   A   N   F   I   N   R   N   S   T   A   E   A   K   V   E   E   I   S   G   V   A   F   I   F   N   Q   K   F   F   Q   D   L   K   E
841/281                                 871/291                                 901/301                                 931/311
gaa aca ggc ata gat ctt gag aac att gtt tac tac aag gac tgc acc cac tat ttt gta atg aca gcc aag aag cag agc ctg ctc gac aaa ggt gtc atc att aac gac tac atc gac
E   T   G   I   D   L   E   N   I   V   Y   Y   K   D   C   T   H   Y   F   V   M   T   A   K   K   Q   S   L   L   D   K   G   V   I   I   N   D   Y   I   D
961/321                                 991/331                                 1021/341                                1051/351
aca gag atg ctg ctg tgt gcg gag aac gtg aac caa gac aac ctg cta tcc tat gcc cgg gaa gct gca gac ttt gcc acc aac tac cag ctg cca tcc tta gac ttt gcc atg aac cac
T   E   M   L   L   C   A   E   N   V   N   Q   D   N   L   L   S   Y   A   R   E   A   A   D   F   A   T   N   Y   Q   L   P   S   L   D   F   A   M   N   H
1081/361                                1111/371                                1141/381                                1171/391
tat ggg cag cct gat gtg gcc atg ttt gac ttt acc tgc atg tat gcc tca gag aac gcg gcc ctg gtg cgg gag cgg cag gcg cac cag ctg ctc gtg gcc ctt gtg ggt gac agc ttg
Y   G   Q   P   D   V   A   M   F   D   F   T   C   M   Y   A   S   E   N   A   A   L   V   R   E   R   Q   A   H   Q   L   L   V   A   L   V   G   D   S   L
1201/401                                1231/411                                1261/421                                1291/431
ctt gag cca ttt tgg ccc atg ggt aca ggc tgt gcc cgt ggc ttc ctg gca gcc ttt gac acg gca tgg atg gtg aag agc tgg aac cag ggc acc cct ccc ctg gag ctg ctg gct gaa
L   E   P   F   W   P   M   G   T   G   C   A   R   G   F   L   A   A   F   D   T   A   W   M   V   K   S   W   N   Q   G   T   P   P   L   E   L   L   A   E
1321/441                                1351/451                                1381/461                                1411/471
agg gaa agt ctc tac cgg ctg tta cct cag aca acc ccg gag aac atc aac aag aac ttt gag cag tac acg ttg gac cca ggg aca cgg tac cca aac ctc aac tca cac tgt gtc agg
R   E   S   L   Y   R   L   L   P   Q   T   T   P   E   N   I   N   K   N   F   E   Q   Y   T   L   D   P   G   T   R   Y   P   N   L   N   S   H   C   V   R
1441/481                                1471/491                                1501/501                                1531/511
ccc cat cag gtg aag cat ttg tat atc act aag gag ctg gag cac tac cct ctc gag aga ctg ggc tcg gtg agg aga tct gtc aac ctc tcc agg aag gag tca gat atc cgg ccc agc
P   H   Q   V   K   H   L   Y   I   T   K   E   L   E   H   Y   P   L   E   R   L   G   S   V   R   R   S   V   N   L   S   R   K   E   S   D   I   R   P   S
1561/521                                1591/531                                1621/541                                1651/551
aag ctc ctg acc tgg tgc cag cag cag aca gag ggc tac cag cat gtc aac gtc acc gac ctg acc aca tcc tgg cgc agt ggg ttg gcc ctg tgt gcc atc atc cac cgc ttc cgg cct
K   L   L   T   W   C   Q   Q   Q   T   E   G   Y   Q   H   V   N   V   T   D   L   T   T   S   W   R   S   G   L   A   L   C   A   I   I   H   R   F   R   P
1681/561                                1711/571                                1741/581                                1771/591
gag ctc atc aac ttt gac tct ttg aat gaa gat gat gct gtg gag aac aac cag ctc gca ttt gat gtg gcc gag cga gag ttt ggg atc cct cca gtg acc acg ggc aaa gag atg gca
E   L   I   N   F   D   S   L   N   E   D   D   A   V   E   N   N   Q   L   A   F   D   V   A   E   R   E   F   G   I   P   P   V   T   T   G   K   E   M   A
1801/601                                1831/611                                1861/621                                1891/631
tct gcc cag gag cct gac aag ctc agc atg gtc atg tac ctc tcc aag ttc tac gag ctc ttc cgg ggc acc cca ctg agg ccc gtg gat tct tgg cgc aaa aac tat gga gaa aat gct
S   A   Q   E   P   D   K   L   S   M   V   M   Y   L   S   K   F   Y   E   L   F   R   G   T   P   L   R   P   V   D   S   W   R   K   N   Y   G   E   N   A
1921/641                                1951/651                                1981/661                                2011/671
gac ctc agc ttg gcc aaa tca tcc att tct aat aac tat ctc aac ctc aca ttt cca agg aag agg act cca cgg gtg gat ggt caa acc gga gag aat gac atg aac aaa cgg aga cgg
D   L   S   L   A   K   S   S   I   S   N   N   Y   L   N   L   T   F   P   R   K   R   T   P   R   V   D   G   Q   T   G   E   N   D   M   N   K   R   R   R
2041/681                                2071/691                                2101/701                                2131/711
aag ggc ttc acc aac ctg gac gag cct tca aac ttt tcc agc cgt agc ttg ggc tcc aat caa gag tgt ggg agc agt aag gaa ggt gga aat cag aac aaa gtc aag tcc atg gcg aat
K   G   F   T   N   L   D   E   P   S   N   F   S   S   R   S   L   G   S   N   Q   E   C   G   S   S   K   E   G   G   N   Q   N   K   V   K   S   M   A   N
```

FIGURE 12-1

```
HMICAL-2.txt
2161/721                                2191/731                                2221/741                                2251/751
cag ctg ctg gcc aag ttt gag gag agc act cgg aac ccc tca ctc atg aag cag gaa cgc cgt gtc tca ggg ata ggt aag ccg gtc ctg tgc tct tcc tcc ggc cct cct gtt cac tct
Q   L   L   A   K   F   E   E   S   T   R   N   P   S   L   M   K   Q   E   R   R   V   S   G   I   G   K   P   V   L   C   S   S   S   G   P   P   V   H   S
2281/761                                2311/771                                2341/781                                2371/791
tgc tgc ccc aag ccg gag gag gcc aca ccc agc cca tca cct cct ctg aaa agg cag ttc ccc tct gtg gtc gtg acg ggg cac gtg ctc aga gag ctc aag caa gtg tct gct ggc agt
C   C   P   K   P   E   E   A   T   P   S   P   S   P   P   L   K   R   Q   F   P   S   V   V   V   T   G   H   V   L   R   E   L   K   Q   V   S   A   G   S
2401/801                                2431/811                                2461/821                                2491/831
gag tgc ctg agc aga cct tgg aga gcc aga gcc aag tct gac cta cag ctg ggt ggg aca gaa aat ttc gct acc ctg cct tct acc cgc ccg agg gcg cag gct ctt tcc ggg gtg ctg
E   C   L   S   R   P   W   R   A   R   A   K   S   D   L   Q   L   G   G   T   E   N   F   A   T   L   P   S   T   R   P   R   A   Q   A   L   S   G   V   L
2521/841                                2551/851                                2581/861                                2611/871
tgg cgg ctg cag caa gtg gag gaa aag att ctc cag aag agg gct cag aac ttg gcc aac agg gaa ttt cac aca aag aac att aag gag aag gcg gct cac ctt gcc tcc atg ttt gga
W   R   L   Q   Q   V   E   E   K   I   L   Q   K   R   A   Q   N   L   A   N   R   E   F   H   T   K   N   I   K   E   K   A   A   H   L   A   S   M   F   G
2641/881                                2671/891                                2701/901                                2731/911
cac ggg gat ttc ccg cag aat aaa ctg ctc tct aaa ggc ctg tct cat act cat cct cca tct cct ccc tct cgc ctt ccg tct cct gat cca gct gct tct tcc tct cca tca act gtt
H   G   D   F   P   Q   N   K   L   L   S   K   G   L   S   H   T   H   P   P   S   P   P   S   R   L   P   S   P   D   P   A   A   S   S   S   P   S   T   V
2761/921                                2791/931                                2821/941                                2851/951
gac tct gct tct cct gcc aga aag gaa aag aag tca cct tca ggg ttc cat ttt cat ccc agc cat ttg aga aca gtg cat cct cag ctg acg gta ggg aaa gtg tcc agc gga ata ggg
D   S   A   S   P   A   R   K   E   K   K   S   P   S   G   F   H   F   H   P   S   H   L   R   T   V   H   P   Q   L   T   V   G   K   V   S   S   G   I   G
2881/961                                2911/971                                2941/981                                2971/991
gct gca gct gaa gtc ctg gtc aat ctg tac atg aat gat cac aga cct aag gcc cag gcc acc tct cca gac ctg gaa tct atg cga aag tca ttt ccc ctt aac ctg gga ggc agc gac
A   A   A   E   V   L   V   N   L   Y   M   N   D   H   R   P   K   A   Q   A   T   S   P   D   L   E   S   M   R   K   S   F   P   L   N   L   G   G   S   D
3001/1001                               3031/1011                               3061/1021                               3091/1031
acg tgt tac ttc tgt aag aaa cgt gtg tac gtg atg gaa cgg ctg agc gcc gag ggc cac ttc ttc cac cgg gag tgt ttc cgc tgc agc atc tgt gcc acc acc ttg cgc ctg gcc gcc
T   C   Y   F   C   K   K   R   V   Y   V   M   E   R   L   S   A   E   G   H   F   F   H   R   E   C   F   R   C   S   I   C   A   T   T   L   R   L   A   A
3121/1041                               3151/1051                               3181/1061                               3211/1071
tac acc ttt gac tgc gat gaa ggc aaa ttt tac tgc aag cct cac ttc att cac tgt aaa acc aat agc aaa caa cgg aag aga cgg gca gag ttg aag caa caa aga gag gag gag gca
Y   T   F   D   C   D   E   G   K   F   Y   C   K   P   H   F   I   H   C   K   T   N   S   K   Q   R   K   R   R   A   E   L   K   Q   Q   R   E   E   E   A
3241/1081                               3271/1091                               3301/1101                               3331/1111
aca tgg caa gag cag gaa gcc cct cgg aga gac act ccc acc gaa agt tct tgc gca gtg gcc gcc att ggc acc ctg gaa ggc agc ccc cca gtt cat ttc agc ctt cca gtg cta cac
T   W   Q   E   Q   E   A   P   R   R   D   T   P   T   E   S   S   C   A   V   A   A   I   G   T   L   E   G   S   P   P   V   H   F   S   L   P   V   L   H
3361/1121
cca ctt ctt ggc.................................................................................................................................................
P   L   L   G
1/1                                     31/11                                   61/21                                   91/31
cac ccc atc tgg ggg aag gac agg agc tgg aca ggc caa gag cta tct ccc ttg gct gga gaa gac cgg gaa aaa ggg agt act gga gcc agg aag gaa gaa gag gga ggg cca gtg ctg
H   P   I   W   G   K   D   R   S   W   T   G   Q   E   L   S   P   L   A   G   E   D   R   E   K   G   S   T   G   A   R   K   E   E   E   G   G   P   V   L
121/41                                  151/51                                  181/61                                  211/71
gta aag gag aag ttg ggc ctg aag aag tta gtc ctc acc cag gag cag aag acc atg ttg ttg gat tgg aat gac tcc atc cct gag agt gtg cac ctc aaa gct ggg gag cga att tcc
V   K   E   K   L   G   L   K   K   L   V   L   T   Q   E   Q   K   T   M   L   L   D   W   N   D   S   I   P   E   S   V   H   L   K   A   G   E   R   I   S
241/81                                  271/91                                  301/101                                 331/111
cag aaa agt gct gag aat ggt aga gga ggc cgt gtg cta aaa cca gtc cgc ccc ctg ctg ctc cct agg gca gca gga gag ccc ctg cca acc cag aga ggg gct cag gag aag atg ggg
Q   K   S   A   E   N   G   R   G   G   R   V   L   K   P   V   R   P   L   L   L   P   R   A   A   G   E   P   L   P   T   Q   R   G   A   Q   E   K   M   G
361/121                                 391/131                                 421/141                                 451/151
acc cct gcg gaa caa gct caa ggg gag cga aac gtg cct cca ccc aag tcc cca ctg cgg ctc ata gcc aat gcc atc cga agg tct cta gag ccc ctc ctt tcc aac tct gaa ggc ggg
T   P   A   E   Q   A   Q   G   E   R   N   V   P   P   P   K   S   P   L   R   L   I   A   N   A   I   R   R   S   L   E   P   L   L   S   N   S   E   G   G
481/161                                 511/171                                 541/181                                 571/191
aag aag gcc tgg gcc aag caa gaa tcc aaa act ttg ccc aca cag gcc tgc act cgc tca ttc agc ctt cgg aaa acc aat tcc aat aaa gac ggg gac cag cat tcc cct ggg aga aac
K   K   A   W   A   K   Q   E   S   K   T   L   P   T   Q   A   C   T   R   S   F   S   L   R   K   T   N   S   N   K   D   G   D   Q   H   S   P   G   R   N
601/201                                 631/211                                 661/221                                 691/231
cag tcc tca gcc ttt agc cct cct gac cct gcc ctc cgc acc cac agt ttg ccc aat cgg cca tcc aag gtc ttt cct gca ctt agg tcc cca ccc tgc agc aag att gaa gaa gtc ccc
Q   S   S   A   F   S   P   P   D   P   A   L   R   T   H   S   L   P   N   R   P   S   K   V   F   P   A   L   R   S   P   P   C   S   K   I   E   E   V   P
721/241                                 751/251                                 781/261                                 811/271
aca ctc ctc gag aaa gtg agt ttg caa gag aac ttc cca gat gct tct aag cct cca aag aaa aga atc tca ctt ttt tcc tcc ctc aga ctc aaa gac aaa tct ttt gag agt ttc ctc
T   L   L   E   K   V   S   L   Q   E   N   F   P   D   A   S   K   P   P   K   K   R   I   S   L   F   S   S   L   R   L   K   D   K   S   F   E   S   F   L
841/281                                 871/291                                 901/301                                 931/311
caa gaa tcc aga caa aga aag gac atc agg gac ctc ttt ggc agc ccc aag agg aag gtg ctg cct gaa gat agt gcg cag gcc ctg gag aag ctg ctg cag cct ttc aaa agc acc tcc
```

FIGURE 12-2

```
                                                       HMICAL-2.txt
Q   E   S   R   Q   R   K   D   I   R   D   L   F   G   S   P   K   R   K   V   L   P   E   D   S   A   Q   A   L   E   K   L   L   Q   P   F   K   S   T   S
961/321                                 991/331                                 1021/341                                1051/351
ctg cgc cag gca gct cct cct cct cct cct cct cct cct cct cct cct cct cct cct cct ccc cct aca gcg gga ggt gca gac tcc aag aac ttt ccc ctc aga gca cag gta aca gag
L   R   Q   A   A   P   P   P   P   P   P   P   P   P   P   P   P   P   P   P   P   P   T   A   G   G   A   D   S   K   N   F   P   L   R   A   Q   V   T   E
1081/361                                1111/371                                1141/381                                1171/391
gct tcc tct tct gcc tct tca acc tcc tcc tcc tct gca gat gaa gaa ttt gat ccc cag ctt tcc ttg cag tta aag gag aag aag aca ctt aga aga aga aag aag cta gaa aaa gca
A   S   S   S   A   S   S   T   S   S   S   S   A   D   E   E   F   D   P   Q   L   S   L   Q   L   K   E   K   K   T   L   R   R   R   K   K   L   E   K   A
1201/401                                1231/411                                1261/421                                1291/431
atg aag cag ttg gtt aag caa gaa gaa ttg aaa aga ctc tat aag gct cag gcc atc cag agg cag ctg gag gag gtg gag gag cgg cag agg gct tct gag atc cag ggt gtg agg ctg
M   K   Q   L   V   K   Q   E   E   L   K   R   L   Y   K   A   Q   A   I   Q   R   Q   L   E   E   V   E   E   R   Q   R   A   S   E   I   Q   G   V   R   L
1321/441                                1351/451                                1381/461                                1411/471
gag aag gcg ttg cga gga gaa gca gat tca ggc aca cag gat gaa gca cag ctt ttg cag gaa tgg ttt aag ctg gtt ctg gag aag aat aaa tta atg cga tat gag tcg gag ctc cta
E   K   A   L   R   G   E   A   D   S   G   T   Q   D   E   A   Q   L   L   Q   E   W   F   K   L   V   L   E   K   N   K   L   M   R   Y   E   S   E   L   L
1441/481                                1471/491                                1501/501                                1531/511
atc atg gcc cag gaa ctg gaa tta gaa gat cat caa agc aga ctg gag cag aaa ctg aga gag aaa atg ctc aag gag gag agc cag aaa gat gag aag gat cta aac gaa gag caa gaa
I   M   A   Q   E   L   E   L   E   D   H   Q   S   R   L   E   Q   K   L   R   E   K   M   L   K   E   E   S   Q   K   D   E   K   D   L   N   E   E   Q   E
1561/521                                1591/531                                1621/541                                1651/551
gta ttc acc gag ctg atg caa gtg att gag caa agg gac aaa ctc gtc gat tcc tta gag gaa caa cgc atc aga gaa aaa gcc gag gac cag cac ttt gaa agc ttc gta ttc tcc aga
V   F   T   E   L   M   Q   V   I   E   Q   R   D   K   L   V   D   S   L   E   E   Q   R   I   R   E   K   A   E   D   Q   H   F   E   S   F   V   F   S   R
1681/561
ggc tgt cag ctg agc agg act tga gga ggc ccg tag tcc ctc tcc ctg gct gca cgt tgg gac cgg atc agg cca agt gca cca cac acc ctc atg ggt ctt tct gca gga ttt atc atc
G   C   Q   L   S   R   T   *
cct gga act taa gtt ata cct tac tgc att ttt taa aat taa aat tct ctt gca cgc atg gca gct tcc caa ggt tct tcc aga gat tca aat gaa gaa aac cca aag act ctt tgg caa
ttg gca gtc aac ttc agc cag gct ctc aga ctg gag gtg ttg ttg gca gat gac cag cat tgt ttt ccc tag aaa gtg aca caa aga ctt gac ttt cct gct act ttt atc att ttc ctt
ccc aat tca ttg agt tac ata ctt taa gat ttt tga gaa gct gcc ttt tca tta ata tat cca tat ttg cct ttt ttt gta tgg atg acc agt ttc caa atg tca gaa aga agc agc cgc
agt tta aag att agg tta ata ttt aaa ttg tgt ttc cag aga aag agg aga aac ctt gag att act gat tac ata aag caa ata act cat ata gca ggt gtt aat tca atc cag ggt gaa
ttt aat tta cca ggt gca ttt ata agc ctt aat ata ata tac ata agc aat gag agc tta ata gaa cat ttg agc ctt aat ttt att tta ag
```

FIGURE 12-3

HMICAL-3.txt

```
1/1                                     31/11                                   61/21                                   91/31
atg gag gag agg aag cat gag acc atg aac cca gct cat gtc ctc ttt gac cgg ttt gtc cag gcc acc acc tgc aag gga acc ctc aag gct ttc cag gag ctc tgt gac cac ctg gaa
M   E   E   R   K   H   E   T   M   N   P   A   H   V   L   F   D   R   F   V   Q   A   T   T   C   K   G   T   L   K   A   F   Q   E   L   C   D   H   L   E
121/41                                  151/51                                  181/61                                  211/71
cta aag cca aag gac tac cgc tcc ttc tat cac aag ctc aag tcc aag ctt aac tac tgg aaa gcc aaa gcc ctc tgg gca aaa ttg gac aaa cgg ggc agt cac aaa gac tac aaa aag
L   K   P   K   D   Y   R   S   F   Y   H   K   L   K   S   K   L   N   Y   W   K   A   K   A   L   W   A   K   L   D   K   R   G   S   H   K   D   Y   K   K
241/81                                  271/91                                  301/101                                 331/111
gga aaa gcg tgc act aac acc aag tgt ctc atc att ggg gct ggc ccc tgt ggt ctc cgt aca gcc atc gac tta tcc tta ctg ggg gcc aag gtg gtt gtt att gag aaa cga gat gcc
G   K   A   C   T   N   T   K   C   L   I   I   G   A   G   P   C   G   L   R   T   A   I   D   L   S   L   L   G   A   K   V   V   V   I   E   K   R   D   A
361/121                                 391/131                                 421/141                                 451/151
ttc tcc cgc aac aac gtc ttg cat ctc tgg cca ttc acc ata cat gat cta cga ggt ctg ggt gcc aag aag ttc tat ggc aag ttc tgt gct gga gcc atc gac cat atc agt atc cgt
F   S   R   N   N   V   L   H   L   W   P   F   T   I   H   D   L   R   G   L   G   A   K   K   F   Y   G   K   F   C   A   G   A   I   D   H   I   S   I   R
481/161                                 511/171                                 541/181                                 571/191
cag ctc caa cta ata ctt ttg aaa gta gcc ttg atc cta ggc att gaa atc cac gtc aat gtg gaa ttc caa gga ctt ata cag cct cct gag gac caa gag aat gaa cgg ata ggc tgg
Q   L   Q   L   I   L   L   K   V   A   L   I   L   G   I   E   I   H   V   N   V   E   F   Q   G   L   I   Q   P   P   E   D   Q   E   N   E   R   I   G   W
601/201                                 631/211                                 661/221                                 691/231
cgg gca ctg gtg cac ccc aag act cat cct gtg tca gag tat gaa ttt gaa gtg atc atc ggt ggg gat ggt cgg agg aac acc ttg gaa ggg ttt cgt cgg aaa gaa ttc cgt ggc aaa
R   A   L   V   H   P   K   T   H   P   V   S   E   Y   E   F   E   V   I   I   G   G   D   G   R   R   N   T   L   E   G   F   R   R   K   E   F   R   G   K
721/241                                 751/251                                 781/261                                 811/271
ctg gcc atc gcc atc acg gca aat ttt atc aac cga aat aca aca gca gaa gct aaa gtg gaa gag atc agt ggt gtg gct ttt ata ttc aac caa aaa ttt ttc cag gaa ctg agg gaa
L   A   I   A   I   T   A   N   F   I   N   R   N   T   T   A   E   A   K   V   E   E   I   S   G   V   A   F   I   F   N   Q   K   F   F   Q   E   L   R   E
841/281                                 871/291                                 901/301                                 931/311
gcc aca ggt att gac ttg gag aac atc gtt tac tac aaa gat gac aca cac tat ttc gtt atg aca gcc aaa aag cag agt ttg ctg gac aaa gga gtg ata cta cat gac tac gcc gac
A   T   G   I   D   L   E   N   I   V   Y   Y   K   D   D   T   H   Y   F   V   M   T   A   K   K   Q   S   L   L   D   K   G   V   I   L   H   D   Y   A   D
961/321                                 991/331                                 1021/341                                1051/351
aca gag ctc ctg ctt tcc cga gaa aac gtg gac cag gag gct ctg ctc agc tat gcc agg gag gcg gca gac ttc tct acc cag cag cag ctg ccg tct ctg gat ttt gcc atc aat cac
T   E   L   L   L   S   R   E   N   V   D   Q   E   A   L   L   S   Y   A   R   E   A   A   D   F   S   T   Q   Q   Q   L   P   S   L   D   F   A   I   N   H
1081/361                                1111/371                                1141/381                                1171/391
tat ggg cag ccc gat gtg gcc atg ttt gac ttc act tgt atg tat gcc tcc gag aac gcc gcc ttg gtg cgg gag cag aac gga cac cag tta cta gtg gct ctg gtc ggg gac agc ctc
Y   G   Q   P   D   V   A   M   F   D   F   T   C   M   Y   A   S   E   N   A   A   L   V   R   E   Q   N   G   H   Q   L   L   V   A   L   V   G   D   S   L
1201/401                                1231/411                                1261/421                                1291/431
cta gag cct ttc tgg cca atg gga aca gga ata gcc cgg ggc ttt cta gct gct atg gac tct gcc tgg atg gtc cga agt tgg tct cta gga acg agc cct ttg gaa gtg ctg gca gag
L   E   P   F   W   P   M   G   T   G   I   A   R   G   F   L   A   A   M   D   S   A   W   M   V   R   S   W   S   L   G   T   S   P   L   E   V   L   A   E
1321/441                                1351/451                                1381/461                                1411/471
agg gaa agt att tac agg ttg ctg cct cag acc acc cct gag aat gtg agt aag aac ttc agc cag tac agt atc gac cct gtc act cgg tat ccc aat atc aac gtc aac ttc ctc cgg
R   E   S   I   Y   R   L   L   P   Q   T   T   P   E   N   V   S   K   N   F   S   Q   Y   S   I   D   P   V   T   R   Y   P   N   I   N   V   N   F   L   R
1441/481                                1471/491                                1501/501                                1531/511
cca agc cag gtg cgc cat tta tat gat act ggc gaa aca aaa gat att cac ctg gaa atg gag agc ctg gtg aat tcc cga acc acc ccc aaa ttg act cgc aat gag tct gta gct cgt
P   S   Q   V   R   H   L   Y   D   T   G   E   T   K   D   I   H   L   E   M   E   S   L   V   N   S   R   T   T   P   K   L   T   R   N   E   S   V   A   R
1561/521                                1591/531                                1621/541                                1651/551
tca agc aaa ctg ctg ggt tgg tgc cag agg cag aca gat ggc tat gca ggg gta aac gtg aca gat ctc acc atg tcc tgg aaa agt ggc ttg gcc ctt tgt gca att atc cat aga tac
S   S   K   L   L   G   W   C   Q   R   Q   T   D   G   Y   A   G   V   N   V   T   D   L   T   M   S   W   K   S   G   L   A   L   C   A   I   I   H   R   Y
1681/561                                1711/571                                1741/581                                1771/591
cgc cct gac ctg ata gat ttt gat tct ttg gat gag caa aat gtg gag aag aat aac caa ctg gcc ttt gac att gct gag aag gaa ttg ggc att tct ccc atc atg aca ggc aaa gaa
R   P   D   L   I   D   F   D   S   L   D   E   Q   N   V   E   K   N   N   Q   L   A   F   D   I   A   E   K   E   L   G   I   S   P   I   M   T   G   K   E
1801/601                                1831/611                                1861/621                                1891/631
atg gcc tcc gtg ggg gag cct gat aag ctg tcc atg gtg atg tac ctg act cag ttc tac gag atg ttt aag gac tcc ctc ccc tct agc gac acc ttg gac cta aat gcc gag gag aaa
M   A   S   V   G   E   P   D   K   L   S   M   V   M   Y   L   T   Q   F   Y   E   M   F   K   D   S   L   P   S   S   D   T   L   D   L   N   A   E   E   K
1921/641                                1951/651                                1981/661                                2011/671
gca gtc ctg ata gcc agc acc aga tcc cct atc tcc ttc cta agc aaa ctt ggc cag acc atc tct cgg aag cgt tct ccc aag gat aaa aag gaa aag gac ttg gat ggt gct ggg aag
A   V   L   I   A   S   T   R   S   P   I   S   F   L   S   K   L   G   Q   T   I   S   R   K   R   S   P   K   D   K   K   E   K   D   L   D   G   A   G   K
2041/681                                2071/691                                2101/701                                2131/711
agg aga aag acc agt caa tca gag gag gag gaa gct cct cgg ggc cac aga gga gaa aga ccg acc ctg gtg agc act ctg aca gac agg agg atg gac gtt gcc gtt ggg aac cag aac
R   R   K   T   S   Q   S   E   E   E   E   A   P   R   G   H   R   G   E   R   P   T   L   V   S   T   L   T   D   R   R   M   D   V   A   V   G   N   Q   N
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 13-1

HMICAL-3.txt

```
aaa gtg aag tac atg gcg acc cag ctg ctg gcc aaa ttt gaa gag aat gcg ccc gca cag tcc atc ggc ata cgg aga cag ctg aca caa gag cgt ggg gcc agc cag ccg tcc tgc tgc
K   V   K   Y   M   A   T   Q   L   L   A   K   F   E   E   N   A   P   A   Q   S   I   G   I   R   R   Q   L   T   Q   E   R   G   A   S   Q   P   S   C   C
2281/761                                2311/771                                2341/781                                2371/791
ctg cct ggg cag gtt cgc cct gcc ccc acc ccc cgg tgg aaa cag ggc tcc atg aag aag gag ttc ccg cag aac ctg gga ggc agc gac aca tgc tac ttc tgc cag aag cgg gtc tac
L   P   G   Q   V   R   P   A   P   T   P   R   W   K   Q   G   S   M   K   K   E   F   P   Q   N   L   G   G   S   D   T   C   Y   F   C   Q   K   R   V   Y
2401/801                                2431/811                                2461/821                                2491/831
gtg atg gag agg ctg agt gcc gag ggc aag ttc ttc cac cgg agc tgc ttc aag tgc gag tac tgc gcc acc acc ctg cgc ctc tcg gcc tac gcc tac gac atc gag gat ggt aaa ttc
V   M   E   R   L   S   A   E   G   K   F   F   H   R   S   C   F   K   C   E   Y   C   A   T   T   L   R   L   S   A   Y   A   Y   D   I   E   D   G   K   F
2521/841                                2551/851                                2581/861                                2611/871
tac tgt aag cca cac tac tgc tat cga ctc tct ggc tac gca caa agg aag aga ccg gca gtg gct ccc ctg tct gga aag gag gcc aaa gga ccc ctg cag gat ggc gcc acc aca gat
Y   C   K   P   H   Y   C   Y   R   L   S   G   Y   A   Q   R   K   R   P   A   V   A   P   L   S   G   K   E   A   K   G   P   L   Q   D   G   A   T   T   D
2641/881                                2671/891                                2701/901                                2731/911
gca aac gga cgg gcc aac gcc gtg gcc agc tcc act gag aga acc cca ggt tca ggc gtg aac ggc ctg gag gag ccc agc atc gcc aag cga ctg agg ggc acc cca gag cgg atc gag
A   N   G   R   A   N   A   V   A   S   S   T   E   R   T   P   G   S   G   V   N   G   L   E   E   P   S   I   A   K   R   L   R   G   T   P   E   R   I   E
2761/921                                2791/931                                2821/941                                2851/951
ctg gag aac tac cgc ctg tcc ctg agg cag gct gag gca ctg cag gag gta ccg gag gag act cag gcc gag cac aac ctg agc agc gtg ctg gac acg ggc gcc gag gag gac gtc gcc
L   E   N   Y   R   L   S   L   R   Q   A   E   A   L   Q   E   V   P   E   E   T   Q   A   E   H   N   L   S   S   V   L   D   T   G   A   E   E   D   V   A
2881/961                                2911/971
agc agg tca gca cgc agg gct gca ggg cgc cca ccc gcc aca cgg ccc.............................................................................................
S   R   S   A   R   R   A   A   G   R   P   P   A   T   R   P
1/1                                     31/11                                   61/21                                   91/31
gaa gag tcc agt gaa gcc ggg aac cag agg ctc cag cag gtc atg cac gcg gcg gat cct ctg gag atc cag gct gac gtg cac tgg act cat atc cgt gag aga gag gag gaa gag agg
E   E   S   S   E   A   G   N   Q   R   L   Q   Q   V   M   H   A   A   D   P   L   E   I   Q   A   D   V   H   W   T   H   I   R   E   R   E   E   E   E   R
121/41                                  151/51                                  181/61                                  211/71
atg gcg ccg gcc tct gag tcc tct gct tcc gga gcc cca ttg gat gag aat gac cta gag gaa gat gtg gac tca gaa cca gcc gag ata gaa ggg gag gca gca gag gat ggg gac cca
M   A   P   A   S   E   S   S   A   S   G   A   P   L   D   E   N   D   L   E   E   D   V   D   S   E   P   A   E   I   E   G   E   A   A   E   D   G   D   P
241/81                                  271/91                                  301/101                                 331/111
ggg gac act ggt gct gag ctg gat gat gat cag cac tgg tct gac agc ccg tcg gat gct gac aga gag ctg cgt ttg ccg tgc cca gct gag ggg gaa gca gag ctg gag ctg agg gtg
G   D   T   G   A   E   L   D   D   D   Q   H   W   S   D   S   P   S   D   A   D   R   E   L   R   L   P   C   P   A   E   G   E   A   E   L   E   L   R   V
361/121                                 391/131                                 421/141                                 451/151
tcg gaa gat gag gag aag ctg ccc gcc tca ccg aag cac caa gag aga ggt ccc tcc caa gcc acc agc ccc atc cgg tct ccc cag gaa tca gct ctt ctg ttc att cca gtc cac agc
S   E   D   E   E   K   L   P   A   S   P   K   H   Q   E   R   G   P   S   Q   A   T   S   P   I   R   S   P   Q   E   S   A   L   L   F   I   P   V   H   S
481/161                                 511/171                                 541/181                                 571/191
ccc tca aca gag ggg ccc caa ctc cca cct gtc cct gcc gcc acc cag gag aaa tca cct gag gag cgc ctt ttc cct gag cct ttg ctc ccc aaa gag aag ccc aaa gct gat gcc ccc
P   S   T   E   G   P   Q   L   P   P   V   P   A   A   T   Q   E   K   S   P   E   E   R   L   F   P   E   P   L   L   P   K   E   K   P   K   A   D   A   P
601/201                                 631/211                                 661/221                                 691/231
tcg gat ctg aaa gct gtg cac tct ccc atc cga tca cag cca gtg acc ctg cca gaa gct agg act cct gtc tca cca ggg agc ccg cag ccc cgg cca ccc gtg gcg gcc tcc acg ccc
S   D   L   K   A   V   H   S   P   I   R   S   Q   P   V   T   L   P   E   A   R   T   P   V   S   P   G   S   P   Q   P   R   P   P   V   A   A   S   T   P
721/241                                 751/251                                 781/261                                 811/271
cca ccc agc cca ctc ccc atc tgc tcc cag ccc cag cct tcc acc gag gcc act gtc cca tcc cct acc cag tcc ccc ata cgc ttc cag cct gcc ccg gcc aaa aca tcc acc cca ctg
P   P   S   P   L   P   I   C   S   Q   P   Q   P   S   T   E   A   T   V   P   S   P   T   Q   S   P   I   R   F   Q   P   A   P   A   K   T   S   T   P   L
841/281                                 871/291                                 901/301                                 931/311
gcc cct ctc cct gtc caa agc caa agt gac acc aag gac aga ctg ggc agc ccc ctt gct gtg gat gag gcc ctc aga cgg agc gac ctg gtg gag gag ttc tgg atg aag agt gcg gag
A   P   L   P   V   Q   S   Q   S   D   T   K   D   R   L   G   S   P   L   A   V   D   E   A   L   R   R   S   D   L   V   E   E   F   W   M   K   S   A   E
961/321                                 991/331                                 1021/341                                1051/351
atc cgc cgc agc ctc ggg ctc aca cct gtg gac cgc agc aag ggg ccc gag ccc agc ttc ccc acg cct gcc ttc agg cca gtg tcc ctc aaa tcc tat tcc gtt gaa aag tcc ccc cag
I   R   R   S   L   G   L   T   P   V   D   R   S   K   G   P   E   P   S   F   P   T   P   A   F   R   P   V   S   L   K   S   Y   S   V   E   K   S   P   Q
1081/361                                1111/371                                1141/381                                1171/391
gat gag gga ctc cac ctt ctc aag cct ctg tcc atc ccc aaa agg ctg ggc ctg cca aag ccg gaa ggc gag ccg ttg tcc ctg cca acc ccc cgg tcc ccg tcc gac aga gag cta cgc
D   E   G   L   H   L   L   K   P   L   S   I   P   K   R   L   G   L   P   K   P   E   G   E   P   L   S   L   P   T   P   R   S   P   S   D   R   E   L   R
1201/401                                1231/411                                1261/421                                1291/431
agc gcc cag gag gag cgc agg gag ctg tcc agc agc tct ggc ctg ggc ctg cac ggg agc tcc tcc aac atg aag aca ctg ggc agc cag agc ttc aac acc tcg gac tcc gcc atg ctc
S   A   Q   E   E   R   R   E   L   S   S   S   S   G   L   G   L   H   G   S   S   S   N   M   K   T   L   G   S   Q   S   F   N   T   S   D   S   A   M   L
1321/441                                1351/451                                1381/461                                1411/471
acg ccc ccc tcc agc ccg ccc cca ccg cca ccc ccg ggc gag gag ccc gcc acc ttg cgg agg aag ctc agg gag gcc gag ccc aat gcc tcg gtg gtc ccg ccg ccc ttg ccc gcc acc
T   P   P   S   S   P   P   P   P   P   P   P   G   E   E   P   A   T   L   R   R   K   L   R   E   A   E   P   N   A   S   V   V   P   P   P   L   P   A   T
```

FIGURE 13-2

HMICAL-3.txt

1441/481 1471/491 1501/501 1531/511
tgg atg cgg ccc ccc cgg gag cct gct cag ccc ccc aga gag gag gtg cgg aag tcg ttt gtg gag agt gtg gag gag att ccc ttt gct gat gat gtg gag gac acc tat gac gac aag
W M R P P R E P A Q P P R E E V R K S F V E S V E E I P F A D D V E D T Y D D K
1561/521 1591/531 1621/541 1651/551
act gag gac tca agc ctg cag gag aaa ttc ttc acg ccc ccg tcc tgc tgg ccg cgc ccc gag aag cct cgc cac ccg ccc ctg gcc aag gag aac ggg agg ctg cct gct ctg gag ggg
T E D S S L Q E K F F T P P S C W P R P E K P R H P P L A K E N G R L P A L E G
1681/561 1711/571 1741/581 1771/591
acg ctg cag cca cag aag agg ggg ctg ccc ttg gtg tcc gcg gaa gcc aag gag ttg gcc gag gag cgc atg cga gcc agg gag aag tcc gtg aag agc cag gcg ctg cgg gat gcc atg
T L Q P Q K R G L P L V S A E A K E L A E E R M R A R E K S V K S Q A L R D A M
1801/601 1831/611 1861/621 1891/631
gcc agg cag ctg agc agg atg cag cag atg gag ctg gcc tca ggc gcc ccc agg ccc cgc aag gcg tcc tca gca ccc tcc cag ggc aag gag cgc cgg cct gac tcc ccc aca cgc ccc
A R Q L S R M Q Q M E L A S G A P R P R K A S S A P S Q G K E R R P D S P T R P
1921/641 1951/651 1981/661 2011/671
act ctc agg ggc tcc gag gag ccc acc ctg aag cat gaa gcc acc agc gag gag gtc ctc tcc ccg ccg tcg gac tca ggg ggc cca gat ggc tct ttc act tca tcc gag ggc tcc agt
T L R G S E E P T L K H E A T S E E V L S P P S D S G G P D G S F T S S -E G S S
2041/681 2071/691 2101/701 2131/711
ggg aag agc aag aag agg tcg tca ctc ttc tcc ccc cgc aga aac aag aag gag aag aag tcc aaa ggc gag ggc cgg ccc ccg gag aag ccc agc tcc aac ctc ctg gaa gaa gcc gcc
G K S K K R S S L F S P R R N K K E K K S K G E G R P P E K P S S N L L E E A A
2161/721 2191/731 2221/741 2251/751
gcc aaa ccc aag tcc ctg tgg aag tcc gtc ttc tcc ggg tac aag aag gac aag aag aag aag gcc gac gac aag tcc tgc ccc agc acc ccc tcc agc ggg gcc acg gtg gac tct gga
A K P K S L W K S V F S G Y K K D K K K K A D D K S C P S T P S S G A T V D S G
2281/761 2311/771 2341/781 2371/791
aag cac agg gtg ctt ccc gtc gta agg gca gag ctg cag ctc cgg cgc cag ctg agc ttc tcc gag gac tca gac ctc tcc agc gac gat gtc ctt gag aag tcc tca cag aag tcc cgg
K H R V L P V V R A E L Q L R R Q L S F S E D S D L S S D D V L E K S S Q K S R
2401/801 2431/811 2461/821 2491/831
cga gag cca aga acc tac acg gag gag gaa ctg aat gcc aag ctg acc cgg cgt gtg caa aag gca gct cgg aga cag gcc aag cag gag gag ctt aag cgg ctg cat cga gcc cag atc
R E P R T Y T E E E L N A K L T R R V Q K A A R R Q A K Q E E L K R L H R A Q I
2521/841 2551/851 2581/861 2611/871
atc cag cgg cag ctg cag cag gtg gag gag agg cag cgg cgg ctg gag gaa agg ggc gtg gct gtg gag aag gcg ctc cgg ggc gaa gca ggc atg ggc aag aag gac gac ccc aag ctg
I Q R Q L Q Q V E E R Q R R L E E R G V A V E K A L R G E A G M G K K D D P K L
2641/881 2671/891 2701/901 2731/911
atg cag gag tgg ttc aag cta gtg cag gag aag aac gcc atg gtg cgc tac gag tcg gag ctg atg atc ttt gcc cgg gag ctg gag ctg gaa gac cgg cag agt cga ctg cag cag gag
M Q E W F K L V Q E K N A M V R Y E S E L M I F A R E L E L E D R Q S R L Q Q E
2761/921 2791/931 2821/941 2851/951
ctc cgg gaa cgc atg gca gtg gaa gat cac ctt aag act gag gag gag ctg tca gaa gag aag cag att ctc aat gag atg ctg gag gtg gtg gag cag aga gac tca ctg gtg gcg ctg
L R E R M A V E D H L K T E E E L S E E K Q I L N E M L E V V E Q R D S L V A L
2881/961 2911/971 2941/981 2971/991
ctg gag gag cag cgg ctc cgg gag aga gag gag gac aag gac ctg gag gct gcc atg ctg tcc aag ggc ttc agc ctt aac tgg tcc tga gct ccc acc caa cgc tcc att ttc tgt tgg
L E E Q R L R E R E E D K D L E A A M L S K G F S L N W S *
cat ctg cct ggc cag gca gtg gca tcc aaa cca ccc gga gcc gcg atc tga gga ggc ctg gca cct cct tgg agt tta cgc tca gat gcc cgt gtg ctg ctt gga aag tgg tcg agt ccc
gcg tgc agt ggg gag ccc cag gtg aca gtg gtt atc tga gac ggc tcc acc tcc tgg gag gag gcc cac ctg gac ctc cca ctc aga gga gga gca cgg cgt gta tgg cat gac gca ggg
gac cac ccc gcg cgc tcc ctg agg atg tgc tgg ctg tgc ccc ttt ttt cca ctg gca cat ttg gta aga gag gga agc tgc tcc ccg tca gaa cca cag tgc gcc gtg cga ggg gca ctg
tct tct tca tgc tcc ctg gag cac cac caa aga aac gta aac aat acc cca cga aag cag ggt cag ggg tca ggg tgc gat cga gac cca gga tgg ggg cgt cca gtc atg ccc acc cca
gca tca cag gag aca tgg agg tgc ggg cag gct cct gaa tta tta tgc aaa tta gga gga cgc agg agg ggt ctg ccc tcc agc cga aca cca cac act gga ccc taa gtg gcc aaa tgc
ctg ggc cgc ttg ctg gct gtg gcc tga ggc ttg tgg gtt gct gca ttt tgc ttg tag ttc aca acc att ttg aca ctg gaa aat gct gac ttt ggg gga cag gat gag gcc cta cat tct
aag ccc cca gtt ggc aga cag gca ttg tcc ctg ttc cac att tat gtc ggg aca gga gat gac ctt ttc ctc cgt gtt ttt cct gtg ttt gca cgt tga aat gaa gct gac aac ctg gca
aga cgc tca gcc gct tca aac cct ttt tgt caa tta act tat ttt tta ata ctt gaa aag aag taa ctt cgt ttg tgt atc ttt act aga gga act gat cac ctg cgc ccg ggt gcg gga
gcc aca gcg gca tct ggt gcg tcc tac gcg acc tgg tcc ggg gct gcc cgg tgc tcc tca cgt gca tct att tat tag cct ttc tct tcg tat cac tgg cct ggc tgg cat cag gga gct
gcc cag aac ccc tgt gtg ggt cct cct cac agc ttt ctg tcc cct cct cca ccc ggt gcc tgc ctc acc ctg gcg cta gac cat ctg gac cac tca tgt gat gag ggt gca ttt ccg ttc
tgt ttt ggg cca ggc caa cag cag agc tgc cac tct cac cct ccc agt gag aat tcc ggc tct gca gaa ctc gcc ctt gtc tca gtt tgg ggg cca ggg cat cac ctt cct ccg cac ata
tgt taa aga agg ttt cag gat ggg ccc tca tcc aca cag gcc aag agg gtg caa ggt ggg acc ctg gaa tca tgt ggc tgg tga gaa ttc cct cct ccc agc cta aga ttc acc cag acg
gaa acg cga gtg ctg cag tgg atg ctg gga tgc agg ctg gtg cct ctc aag cag atc agc gac tcc ccc ctt cct tcc gaa ggt gac ggg cac ctg cct tgt gcc gga tct tca tgg gga
cat aaa ggg cga gcc ccg aat cac tag ctc cta tag cca aac tgt tcc ttt ctc acg gtt ccg cac cag cct ggc tgt gta cag ctc atc aag cca ctg agc taa tcg ggg tgg ggt gct
tgt cca tca aag cag tca gca cat agc cca gcg agg agc cat ccg gac cag acc cgc ctg cca ggg gcg ctc cag ctg cct gcc gcc tcc gtg ggc tgg ggc cag ctg gga gca gag gcc
tgg ccc ctc caa ggc gtc ccg cag tgg agc agg ctg agt ggc tgt gct gac ctt ggg ctt tcc atg gga acc aca ctg tgc ttc aac ttg aac att cat ccc agc tgc aaa gga gca aag
aac ctg tgt cat cct tgt ctg tcc aga agc tgc cat ctc tct ccc atg cac atc cgg gag ata acg ccg cta gtg gcc gcc aca gcg ctg att ctc cac ctg ttc tca ctt gag gca aag

FIGURE 13-3

HMICAL-3.txt

```
gtc tcc ttt tct ttt ctc ttt act taa aaa agg gga gag gag ggt ttc tag att cca tct tca aac ccc agt cgt ccc ata aaa ttg gac gtg gga aaa gac ctt cac tgc ctg cgt ggc
ctt tcc cag acc tct cct cga atg cca caa aac cgg tcc agc ccc ggc aga gcc gcc ttc ggc cct tgt agc tcc tgc tgg ccc cac aac agg gaa aca gtt tgc aaa gtg gct tgg aag
gag tgt ggc cta ggg atc tgc tgt aag tgg cca gac gta gca gga gag ccc agt gtc acc ttc tgg tcc tgg tca gcc tta aca ctg tgg cat cct cca cag cac aca tgg cag cct gca
gtc tgg aag gtg gac ccg agc ctt tgc aga gag gcg ccg cag agc cgc agg cct gcg ccc cag cct tct gct ccc gac tgt gag gac acc cag ttc tag tag agc act ttt ttt aaa gct
ccc att ttg taa cca cta gtt tgc ggt tga ctt gag tac tct ggt gac ttc ctg cgt caa gcg ttc tca agc tgt gag aat gtg cgc agc tcc agg cag gtt ttc tct cgg aga gtt aag
tct tcc ctt gaa ggc agg gaa gca gga tgg ata cac ata tat cac aca cat aaa aca cca ggt gcg gga gca gcc cag act caa ggc tga cta aac tgg agg ctg aat acc gtg gag gtc
cac atg cag ctt ccc tgg agg gca ggc cgg agg cgc tcc cgc ccc tgg gct tga gga tgc tgc acc ccg tgg gct tcc agg cct gcc cag atg atg cct tca ggc ctc tgt ccc tgg cgg
cca tcc tca ggc cga ttt tga cca gca atg ata gac tct tct taa ccc ttt caa aat aaa ttt ttc agt ggg aca gaa agg aga gtt aaa aaa cat ttt ttt aaa ggt ggt aac atc tga
ccc aca aag gga atg ggt ctg ttt tat gca aaa taa aag ttt ttc aaa t
```

FIGURE 13-4

```
DMICALlong.txt
GCTCTGCATCTCGCAACGAATCGTTCCGAATAAAAAATTAAATAAATAAACACAAAAAATTGTGATATAAGCGATAAATATTGTGATAATTCGGACAAACGTGTCAATTGCATTGCATCAGA
1/1                                     31/11                                   61/21                                   91/31
ATG AGC CGC CAA CAC CAG CGG CAC CAC CAG CAG CAT CAC CAC CTG CCG CCG CAC CAG CAA CCG CAG CAG CAG ATG CCG CAA CAA CAG CAG CAG CTG ACG GCG CAG CAG CAG CAA CAA CAG
M   S   R   Q   H   Q   R   H   H   Q   Q   H   H   H   L   P   P   H   Q   Q   P   Q   Q   Q   M   P   Q   Q   Q   Q   Q   L   T   A   Q   Q   Q   Q   Q   Q
121/41                                  151/51                                  181/61                                  211/71
CAG CTG CTG ATG GCG GAG CAC GCG GCG GCC GCG GAG GCG GCG GAG CTA TTC GAC CTG CTG TGC GTG GCC ACA ACG ATG CGC CAG ATC CTG GCG CTC CAT CGG GCC ATG TGC GAG GCT GTG
Q   L   L   M   A   E   H   A   A   A   A   E   A   A   E   L   F   D   L   L   C   V   A   T   T   M   R   Q   I   L   A   L   H   R   A   M   C   E   A   V
241/81                                  271/91                                  301/101                                 331/111
GGA TTG AGA CCC TCG CCT CTG AAC GAC TTC TAC CCA CGG CTA AAG GCC AAG GTG CGT TCG TGG AAG GCG CAG GCC CTG TGG AAG AAG TTC GAC GCC AGA GCT GCC CAT AGA GTC TAC GGC
G   L   R   P   S   P   L   N   D   F   Y   P   R   L   K   A   K   V   R   S   W   K   A   Q   A   L   W   K   K   F   D   A   R   A   A   H   R   V   Y   G
361/121                                 391/131                                 421/141                                 451/151
AAG GGA GCT GCC TGT ACT GGC ACA CGC GTC CTG GTC ATC GGA GCA GGG CCC TGT GGA CTG CGC ACC GCC ATC GAG GCC CAA CTG CTG GGC GCC AAG GTG GTG GTG CTG GAG AAA CGC GAT
K   G   A   A   C   T   G   T   R   V   L   V   I   G   A   G   P   C   G   L   R   T   A   I   E   A   Q   L   L   G   A   K   V   V   V   L   E   K   R   D
481/161                                 511/171                                 541/181                                 571/191
CGC ATC ACC CGG AAC AAT GTG CTC CAT CTG TGG CCA TTC GTC ATC ACG GAT CTG CGC AAC TTG GGC GCA AAG AAG TTC TAC GGC AAG TTT TGC GCC GGC TCC ATC GAT CAC ATC TCC ATT
R   I   T   R   N   N   V   L   H   L   W   P   F   V   I   T   D   L   R   N   L   G   A   K   K   F   Y   G   K   F   C   A   G   S   I   D   H   I   S   I
601/201                                 631/211                                 661/221                                 691/231
CGG CAG CTG CAG TGC ATG CTG CTC AAG GTG GCG CTG CTC CTG GGC GTA GAG ATC CAC GAG GGA GTC AGT TTT GAT CAC GCT GTA GAG CCC TCT GGC GAT GGC GGC GGA TGG AGG GCA GCT
R   Q   L   Q   C   M   L   L   K   V   A   L   L   L   G   V   E   I   H   E   G   V   S   F   D   H   A   V   E   P   S   G   D   G   G   G   W   R   A   A
721/241                                 751/251                                 781/261                                 811/271
GTT ACT CCC GCA GAT CAT CCT GTA TCT CAC TAC GAA TTC GAT GTG TTG ATC GGA GCG GAT GGC AAG CGG AAT ATG CTG GAC TTT AGG AGG AAG GAG TTC CGC GGG AAG CTG GCC ATC GCT
V   T   P   A   D   H   P   V   S   H   Y   E   F   D   V   L   I   G   A   D   G   K   R   N   M   L   D   F   R   R   K   E   F   R   G   K   L   A   I   A
841/281                                 871/291                                 901/301                                 931/311
ATT ACA GCG AAC TTT ATC AAC AAG AAG ACG GAG GCG GAG GCT AAA GTA GAG GAG ATC AGT GGG GTG GCT TTC ATC TTC AAC CAG GCC TTC TTC AAG GAG CTG TAC GGG AAG ACG GGC ATC
I   T   A   N   F   I   N   K   K   T   E   A   E   A   K   V   E   E   I   S   G   V   A   F   I   F   N   Q   A   F   F   K   E   L   Y   G   K   T   G   I
961/321                                 991/331                                 1021/341                                1051/351
GAC CTG GAA AAC ATC GTC TAC TAC AAG GAC GAG ACG CAC TAC TTC GTG ATG ACG GCC AAG AAG CAC AGT CTA ATT GAC AAG GGC GTT ATT ATC GAG GAT ATG GCC GAT CCC GGC GAG CTT
D   L   E   N   I   V   Y   Y   K   D   E   T   H   Y   F   V   M   T   A   K   K   H   S   L   I   D   K   G   V   I   I   E   D   M   A   D   P   G   E   L
1081/361                                1111/371                                1141/381                                1171/391
CTC GCC CCA GCC AAT GTG GAT ACA CAA AAG CTG CAC GAC TAT GCA CGC GAG GCT GCG GAG TTC TCC ACC CAA TAC CAA ATG CCA AAC CTG GAG TTC GCT GTT AAT CAC TAC GGC AAA CCA
L   A   P   A   N   V   D   T   Q   K   L   H   D   Y   A   R   E   A   A   E   F   S   T   Q   Y   Q   M   P   N   L   E   F   A   V   N   H   Y   G   K   P
1201/401                                1231/411                                1261/421                                1291/431
GAT GTG GCC ATG TTC GAC TTC ACA TCG ATG TTT GCC GCC GAG ATG TCC TGT CGG GTG ATT GTG CGC AAA GGA GCT CGC CTG ATG CAG TGC CTC GTG GGT GAC AGT CTG CTC GAG CCG TTT
D   V   A   M   F   D   F   T   S   M   F   A   A   E   M   S   C   R   V   I   V   R   K   G   A   R   L   M   Q   C   L   V   G   D   S   L   L   E   P   F
1321/441                                1351/451                                1381/461                                1411/471
TGG CCC ACT GGA TCG GGT TGT GCC CGT GGA TTC TTA TCC AGC ATG GAT GCT GCC TAT GCC ATC AAG CTT TGG TCC AAC CCG CAG AAC AGC ACA CTT GGC GTT CTG GCG CAG CGC GAA AGC
W   P   T   G   S   G   C   A   R   G   F   L   S   S   M   D   A   A   Y   A   I   K   L   W   S   N   P   Q   N   S   T   L   G   V   L   A   Q   R   E   S
1441/481                                1471/491                                1501/501                                1531/511
ATC TAC CGG CTG CTT AAC CAG ACC ACG CCG GAC ACC CTG CAG CGG GAC ATC AGT GCC TAT ACC GTG GAT CCG GCC ACG CGC TAT CCG AAT CTG AAC AGG GAG TCG GTC AAT AGC TGG CAG
I   Y   R   L   L   N   Q   T   T   P   D   T   L   Q   R   D   I   S   A   Y   T   V   D   P   A   T   R   Y   P   N   L   N   R   E   S   V   N   S   W   Q
1561/521                                1591/531                                1621/541                                1651/551
GTC AAA CAT CTG GTC GAC ACG GAC GAC CCG TCC ATT CTG GAG CAG ACC TTC ATG GAC ACG CAT GCT CTG CAG ACC CCG CAT TTG GAC ACA CCG GGC AGA CGC AAG CGA CGC AGT GGA GAC
V   K   H   L   V   D   T   D   D   P   S   I   L   E   Q   T   F   M   D   T   H   A   L   Q   T   P   H   L   D   T   P   G   R   R   K   R   R   S   G   D
1681/561                                1711/571                                1741/581                                1771/591
TTG CTG CCC CAG GGT GCC ACG TTG CTG AGA TGG ATA AGT GCC CAG CTG CAT TCC TAT CAG TTT ATT CCC GAA CTC AAG GAG GCT TCG GAT GTG TTC CGG AAT GGA CGC GTT CTG TGT GCG
L   L   P   Q   G   A   T   L   L   R   W   I   S   A   Q   L   H   S   Y   Q   F   I   P   E   L   K   E   A   S   D   V   F   R   N   G   R   V   L   C   A
1801/601                                1831/611                                1861/621                                1891/631
CTT ATC AAT CGC TAT CGT CCT GAT CTC ATC GAC TAC GCT GCC ACC AAG GAC ATG AGT CCC GTG GAG TGC AAT GAG CTG TCA TTC GCC GTC CTA GAG CGC GAA CTC CAC ATC GAT CGC GTC
L   I   N   R   Y   R   P   D   L   I   D   Y   A   A   T   K   D   M   S   P   V   E   C   N   E   L   S   F   A   V   L   E   R   E   L   H   I   D   R   V
1921/641                                1951/651                                1981/661                                2011/671
ATG AGT GCC AAA CAG TCG CTG GAC TTG ACC GAG CTG GAG TCG CGA ATC TGG CTC AAC TAT TTG GAC CAG ATC TGC GAC TTG TTT CGC GGC GAG ATC CCC CAT ATC AAG CAC CCC AAG ATG
M   S   A   K   Q   S   L   D   L   T   E   L   E   S   R   I   W   L   N   Y   L   D   Q   I   C   D   L   F   R   G   E   I   P   H   I   K   H   P   K   M
2041/681                                2071/691                                2101/701                                2131/711
GAC TTT AGC GAT TTG CGC CAG AAG TAT CGT ATC AAC CAT ACG CAT GCC CAA CCC GAC TTC TCC AAG CTG CTG GCA ACG AAA CCC AAG GCC AAG TCG CCG ATG CAG GAT GCT GTG GAC ATA
D   F   S   D   L   R   Q   K   Y   R   I   N   H   T   H   A   Q   P   D   F   S   K   L   L   A   T   K   P   K   A   K   S   P   M   Q   D   A   V   D   I
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 14-1

```
                                                                   DMICALlong.txt
CCC ACG ACA GTG CAG CGG CGC TCG GTG CTC GAG GAG GAG CGA GCC AAG CGG CAG CGT CGC CAC GAG CAG CTT CTT AAC ATC GGT GGA GGG GCA GCA GGA GCC GCC GCC GGA GTT GCC GGA
P   T   T   V   Q   R   R   S   V   L   E   E   E   R   A   K   R   Q   R   R   H   E   Q   L   L   N   I   G   G   G   A   A   G   A   A   A   G   V   A   G
2281/761                                2311/771                                2341/781                                2371/791
AGC GGG ACA GGA ACC ACA ACG CAG GGT CAA AAC GAT ACG CCA CGC CGG TCC AAG AAG CGC CGT CAG GTT GAC AAA ACC GCC AAT ATT GAG GAG CGC CAG CAG CGC TTG CAG GAG ATC GAG
S   G   T   G   T   T   T   Q   G   Q   N   D   T   P   R   R   S   K   K   R   R   Q   V   D   K   T   A   N   I   E   E   R   Q   Q   R   L   Q   E   I   E
2401/801                                2431/811                                2461/821                                2491/831
GAG AAT CGG CAG GAG CGG ATG AGC AAG CGG CGC CAG CAG CGC TGT CAC CAG ACG CAG AAT TTC TAC AAG AGC CTT CAG CTC CTG CAG GCG GGC AAG CTC TTG AGG GAG GGT GGT GAG GCG
E   N   R   Q   E   R   M   S   K   R   R   Q   Q   R   C   H   Q   T   Q   N   F   Y   K   S   L   Q   L   L   Q   A   G   K   L   L   R   E   G   G   E   A
2521/841                                2551/851                                2581/861                                2611/871
GGA GTG GCC GAG GAT GGC ACC CCA TTC GAG GAC TAC TCG ATA TTC CTC TAC CGC CAG CAG GCG CCC GTA TTC AAT GAT CGC GTC AAG GAC CTC GAG CGA AAG CTT CTG TTT CCC GAT CGC
G   V   A   E   D   G   T   P   F   E   D   Y   S   I   F   L   Y   R   Q   Q   A   P   V   F   N   D   R   V   K   D   L   E   R   K   L   L   F   P   D   R
2641/881                                2671/891                                2701/901                                2731/911
GAA CGG GGA GAT ATT CCG TCG GCA TTG CCG CGC ACG GCG GAC GAG CAG TTC AGC GAT CGG ATT AAA AAC ATG GAG CAG CGG ATG ACG GGA CGT GGT GGC CTT GGT GGT GAC AAG AAG CCC
E   R   G   D   I   P   S   A   L   P   R   T   A   D   E   Q   F   S   D   R   I   K   N   M   E   Q   R   M   T   G   R   G   G   L   G   G   D   K   K   P
2761/921                                2791/931                                2821/941                                2851/951
AAG GAT CTG ATG CGG GCT ATC GGC AAG ATC GAC TCG AAC GAT TGG AAT GTA CGC GAG ATC GAG AAG AAG ATC GAG CTG TCG AAG AAG ACG GAG ATC CAC GGG CCT AAG GGC CGC GAG AAG
K   D   L   M   R   A   I   G   K   I   D   S   N   D   W   N   V   R   E   I   E   K   K   I   E   L   S   K   K   T   E   I   H   G   P   K   G   R   E   K
2881/961                                2911/971                                2941/981                                2971/991
GTT CCC AAG TGG AGC AAG GAG CAG TTC CAG GCG CGA CAG CAC AAG ATG TCC AAG CCG CAA CGC CAG GAT TCG CGT GAG GCG GAA AAG TTC AAG GAC ATC GAC CAG ACT ATC CGC AAC TTG
V   P   K   W   S   K   E   Q   F   Q   A   R   Q   H   K   M   S   K   P   Q   R   Q   D   S   R   E   A   E   K   F   K   D   I   D   Q   T   I   R   N   L
3001/1001                               3031/1011                               3061/1021                               3091/1031
GAC AAG CAG CTA AAG GAG GGC CAC AAT CTG GAT GTG GGC GAG CGG GGA CGC AAC AAA GTG GCC TCC ATT GCC GGT CAG TTT GGC AAA AAG GAT GAG GCC AAT TCG GAT GAG AAG AAC GCG
D   K   Q   L   K   E   G   H   N   L   D   V   G   E   R   G   R   N   K   V   A   S   I   A   G   Q   F   G   K   K   D   E   A   N   S   D   E   K   N   A
3121/1041                               3151/1051                               3181/1061                               3211/1071
GGC AGC AGC AAT GCC ACC ACC AAC ACC AAC AAC ACA GTC ATA CCC AAA TCT AGT TCC AAG GTG GCG CTG GCC TTT AAA AAG CAG GCT GCC TCC GAA AAG TGC CGC TTC TGT AAG CAA ACC
G   S   S   N   A   T   T   N   T   N   N   T   V   I   P   K   S   S   S   K   V   A   L   A   F   K   K   Q   A   A   S   E   K   C   R   F   C   K   Q   T
3241/1081                               3271/1091                               3301/1101                               3331/1111
GTC TAC CCG ATG GAG AAG ACC ACC GTG GAG GGA TTG GTT CTG CAT CGC AAT TGC CTT AAG TGC CAC CAC TGC CAC ACC AAC TTG CGT CTG GGA GGC TAC GCC TTT GAT CGG GAC GAT CCG
V   Y   P   M   E   K   T   T   V   E   G   L   V   L   H   R   N   C   L   K   C   H   H   C   H   T   N   L   R   L   G   G   Y   A   F   D   R   D   D   P
3361/1121                               3391/1131                               3421/1141                               3451/1151
CAG GGC CGA CTT TAC TGC ACC CAA CAC TTC CGG TTG CCA CCC AAA CCG CTG CCG CAG CGC ACC AAC AAA GCC AGG AAA TCC GCT GCC GCT CAA CCC GCC TCG CCT GCT GTA CCA CCA ACT
Q   G   R   L   Y   C   T   Q   H   F   R   L   P   P   K   P   L   P   Q   R   T   N   K   A   R   K   S   A   A   A   Q   P   A   S   P   A   V   P   P   T
3481/1161                               3511/1171                               3541/1181                               3571/1191
GCG GGA TCC GTA CCC ACT GCA GCT GCC ACA TCG GAG CAT ATG GAC ACC ACT CCA CCC AGG GAC CAG GTG GAT CTA CTG CAG ACC TCG CGA GCA AAT GCC TCT GCC GAT GCC ATG TCC GAT
A   G   S   V   P   T   A   A   A   T   S   E   H   M   D   T   T   P   P   R   D   Q   V   D   L   L   Q   T   S   R   A   N   A   S   A   D   A   M   S   D
3601/1201                               3631/1211                               3661/1221                               3691/1231
GAT GAA GCC AAT GTT ATC GAT GAG CAC GAA TGG TCT GGT CGC AAC TTC TTG CCC GAG TCC AAC AAC GAT TCC CAA TCA GAG CTA TCC AGT TCA GAT GAA TCG GAT ACG GAA TCG GAT TCG
D   E   A   N   V   I   D   E   H   E   W   S   G   R   N   F   L   P   E   S   N   N   D   S   Q   S   E   L   S   S   S   D   E   S   D   T   E   S   D   S
3721/1241                               3751/1251                               3781/1261                               3811/1271
GAG ATG TTT GAG GAG GCG GAT GAT TCG CCG TTT GGT GCT CAG ACC CTC CAG CTG GCG TCG GAT TGG ATT GGA AAG CAA TAC TGT GAG GAC AGT GAT GAT TCT GAC GAT TTC TAC GAC TCA
E   M   F   E   E   A   D   D   S   P   F   G   A   Q   T   L   Q   L   A   S   D   W   I   G   K   Q   Y   C   E   D   S   D   D   S   D   D   F   Y   D   S
3841/1281                               3871/1291                               3901/1301                               3931/1311
AGT GAA GAT GAC GGC AAA GAT GAC ACC GAG GGT GAG GAA TTC AAG AAG GCC CGC GAA TTG AGG CGC CAG GAA GTT CGC CTG CAG CCG TTG CCC GCC AAT CTG CCC ACA GAT ACG GAG ACC
S   E   D   D   G   K   D   D   T   E   G   E   E   F   K   K   A   R   E   L   R   R   Q   E   V   R   L   Q   P   L   P   A   N   L   P   T   D   T   E   T
3961/1321                               3991/1331                               4021/1341                               4051/1351
GAG AAA CTC AAG TTA AAC GTA GAC AAT AAA GAA AAC ATG GCG GAC GGA AGC TCT CTG AAA TCG GGC AAT TCC TTT GAG TCC GCG CGC AGC CAG CCG TCT ACG CCC CTG TCC ACG CCC ACT
E   K   L   K   L   N   V   D   N   K   E   N   M   A   D   G   S   S   L   K   S   G   N   S   F   E   S   A   R   S   Q   P   S   T   P   L   S   T   P   T
4081/1361                               4111/1371                               4141/1381                               4171/1391
CGT GTC GAG ATG GAG CAA CTG GAG CGA GAT GCT CCA CGC AAG TTT AGC AGC GAA ATC GAG GCA ATC AGC GAG AAA CTT TAC CAC ATG AAC AAC ATG GTA AAG ATG AAC AAG GAC CTC GAG
R   V   E   M   E   Q   L   E   R   D   A   P   R   K   F   S   S   E   I   E   A   I   S   E   K   L   Y   H   M   N   N   M   V   K   M   N   K   D   L   E
4201/1401                               4231/1411                               4261/1421                               4291/1431
GTG CTG GCC AAG GAA AAT CTG GTC AAG AGC GGC ATC CTG CGC AAG CTC ACG CTG AAG GAG AAG TGG CTG GCG GAG AAT GCG GCC ATA GCA GCA GGC CAA AAA GTG ACT CCG ACT CCT AGT
V   L   A   K   E   N   L   V   K   S   G   I   L   R   K   L   T   L   K   E   K   W   L   A   E   N   A   A   I   A   A   G   Q   K   V   T   P   T   P   S
4321/1441                               4351/1451                               4381/1461                               4411/1471
GCG ACT GCT CCT GGG CTT CAA CCC AAG TCG AAG TTC GAT GAA AAG TTC GAA AAG GTA GTG AGT CCA CCT CAG CCG GTT GTC GAA CCC AAA CCC AAG CCC GTA ATC GAT TTC AAT TTG GAT
A   T   A   P   G   L   Q   P   K   S   K   F   D   E   K   F   E   K   V   V   S   P   P   Q   P   V   V   E   P   K   P   K   P   V   I   D   F   N   L   D
```

FIGURE 14-2

DMICALlong.txt

```
4441/1481                                  4471/1491                                  4501/1501                                  4531/1511
GAG TTA AAA CCA CGT AAA CCG AAC TTT GAG GAG CGA CCC AAG GAG CAG CTG CCC AGG CCT GAA AGT TTG AAA AAA CCT CCA CAA CAA AAA CCG AAG GGC AGC AGC ACA AAC GTG AGC CGC
E   L   K   P   R   K   P   N   F   E   E   R   P   K   E   Q   L   P   R   P   E   S   L   K   K   P   P   Q   Q   K   P   K   G   S   S   T   N   V   S   R
4561/1521                                  4591/1531                                  4621/1541                                  4651/1551
TCC AAC AGC TTG AAG AGT AAT GCC AGC AAC GGG AGT CCA AAG GTG AAA AAA GCT CCT GTT TCA AAT AAC AGC AAA ATG CAA ATT GAA GGA ATC CTT GAT ACC TTG AGA AAG ATT CAG AGC
S   N   S   L   K   S   N   A   S   N   G   S   P   K   V   K   K   A   P   V   S   N   N   S   K   M   Q   I   E   G   I   L   D   T   L   R   K   I   Q   S
4681/1561                                  4711/1571                                  4741/1581                                  4771/1591
CAG AAC AGT AGC GAT CAG GAT GAA GAT ATG GAT GTA GAT GAG GAT GTG GAA AGA AAG CCA AAT AAA GAG CTC AAC AGC AAG TTG AAG GAA ATC CAA GCC AGC AGC TTT GCC GGC ACA ATG
Q   N   S   S   D   Q   D   E   D   M   D   V   D   E   D   V   E   R   K   P   N   K   E   L   N   S   K   L   K   E   I   Q   A   S   S   F   A   G   T   M
4801/1601                                  4831/1611                                  4861/1621                                  4891/1631
GAC CAT ATT AAA TCC CAG CTG ACT ATG CCA ACG GTT AGT GCG CAG GCA CCA CCC TCC ATG GAT CTT TCC AAG TAC TTT CCC AAC CAG AAG CAA GAG AAG AGC TCC ACC AGC AGC ACA AAC
D   H   I   K   S   Q   L   T   M   P   T   V   S   A   Q   A   P   P   S   M   D   L   S   K   Y   F   P   N   Q   K   Q   E   K   S   S   T   S   S   T   N
4921/1641                                  4951/1651                                  4981/1661                                  5011/1671
AAG AAT CAG GTT ACC CTG AAG GAT GTA AAT CTG GCC AAG TAC TTC CCC AGC AGT CCA GCT CCA CAA CGG AGA ACC GTG GAA ACG GTG GCC GAT CGA CTA AAG AAG TCA CAG ACT GAG GCT
K   N   Q   V   T   L   K   D   V   N   L   A   K   Y   F   P   S   S   P   A   P   Q   R   R   T   V   E   T   V   A   D   R   L   K   K   S   Q   T   E   A
5041/1681                                  5071/1691                                  5101/1701                                  5131/1711
GCT CTA GCT AAA ACC AAG CTA CTG GAG GAT CAG GCC AAT AAC CAG GCC GAA AAG ACC AAG AAA GAG GTC GAA AAG GAA GGA GAG TCT AAA AAA ATC ACG AAA AAG GTG GCA GAC TCC AAA
A   L   A   K   T   K   L   L   E   D   Q   A   N   N   Q   A   E   K   T   K   K   E   V   E   K   E   G   E   S   K   K   I   T   K   K   V   A   D   S   K
5161/1721                                  5191/1731                                  5221/1741                                  5251/1751
GCA GTA CCT CCA AAA CGG CAG GCC TCG TTA GAC ACC TTT AGC TTA AGG GAG CAC CAA ATG GAT GGA GCA TTA GAC CTG ACC AAA AAG AAG GGT CCC ACC AAA GCG AGC GCT GGT GTT AAA
A   V   P   P   K   R   Q   A   S   L   D   T   F   S   L   R   E   H   Q   M   D   G   A   L   D   L   T   K   K   K   G   P   T   K   A   S   A   G   V   K
5281/1761                                  5311/1771                                  5341/1781                                  5371/1791
AAA CCA GCC AAA TCA GGG AGT ACC ACA TCC GTG ACA AAA GCC ACG GCC ACA TCC GAA GGA AAG ACC ATT AAG ATT GTA AAG AAA ATC GTA CCC AAA GGA ACC AAA GCT AAA AAG GCA GCA
K   P   A   K   S   G   S   T   T   S   V   T   K   A   T   A   T   S   E   G   K   T   I   K   I   V   K   K   I   V   P   K   G   T   K   A   K   K   A   A
5401/1801                                  5431/1811                                  5461/1821                                  5491/1831
GAA GCG GCC CAA GAA TCC GCA GTA GTA GAG GCT CCG CCA GAA AAG AAA CCT CCA AAA GAT GAA GCA GAG CGG ATA TTG GAC GAA ATT CTG GGA GAT GGT GAG TAT CGT TCA CCC AGC TCC
E   A   A   Q   E   S   A   V   V   E   A   P   P   E   K   K   P   P   K   D   E   A   E   R   I   L   D   E   I   L   G   D   G   E   Y   R   S   P   S   S
5521/1841                                  5551/1851                                  5581/1861                                  5611/1871
GAG TAT CAG CGC TTG TTC CAG GAC GAA AAG TCG CCT AGC GAT TTG TCC GAT AAC ATA GAT CGG ATA CTC GAG GAA TCC GAG TTG GAT GTG GAA TTG GGT CTG CCA AAA CGC AGC AGT AAG
E   Y   Q   R   L   F   Q   D   E   K   S   P   S   D   L   S   D   N   I   D   R   I   L   E   E   S   E   L   D   V   E   L   G   L   P   K   R   S   S   K
5641/1881                                  5671/1891                                  5701/1901                                  5731/1911
AAA CTG GTA AAA ACC AAA TCC CTG GGC GAG GGT GAC TTT GAC ATG AAG CCT TCA AAA GAA AGA CTT ACT GGA GTA CAG AAC ATA CTT AAG CGA TTT GAG TCA ATG AGT TCA GTC ACC TCA
K   L   V   K   T   K   S   L   G   E   G   D   F   D   M   K   P   S   K   E   R   L   T   G   V   Q   N   I   L   K   R   F   E   S   M   S   S   V   T   S
5761/1921                                  5791/1931                                  5821/1941                                  5851/1951
CAG AAC AGT GAT GAA CAG GCG GGG TTC AAA TTA CGT CGC ATG GAG TCC ACC ACC AGT AAT CTG AGC AGC TTG ACC CGT TCC AGG GAA TCA CTT GTC TCC GTC AGC GAT TCA ATG AGT GAT
Q   N   S   D   E   Q   A   G   F   K   L   R   R   M   E   S   T   T   S   N   L   S   S   L   T   R   S   R   E   S   L   V   S   V   S   D   S   M   S   D
5881/1961                                  5911/1971                                  5941/1981                                  5971/1991
TTG GAG AAA ACC ATG GAC TAC CTC CGT AAT GAA TGG CGC AAT GAG GCC ACC AAT TTT CTG CAA AAG AAG CGT GAT AAA TTC TAT GCC AAA AAG GAG GAG CAA GAA AAG GAG GCT AAA ATT
L   E   K   T   M   D   Y   L   R   N   E   W   R   N   E   A   T   N   F   L   Q   K   K   R   D   K   F   Y   A   K   K   E   E   Q   E   K   E   A   K   I
6001/2001                                  6031/2011                                  6061/2021                                  6091/2031
CTG GCT AAA CCT GAT CCC TTA GAC AAT CTT CCT GTT CAA TAT CGC GAC TCC AAG TTG GCC AAG TTC TTT GGC TTG GCC GCT AGC AAG TCA CCA GAA AAC CGA AAG TCC CCT ATT AAG AAG
L   A   K   P   D   P   L   D   N   L   P   V   Q   Y   R   D   S   K   L   A   K   F   F   G   L   A   A   S   K   S   P   E   N   R   K   S   P   I   K   K
6121/2041                                  6151/2051                                  6181/2061                                  6211/2071
AAG AAA TCT CCC TCT AAA ACG CCA AAG GTG ACC AAA GCA AAC AAT TCA CTG GAG GAG CTG GCC AAG ATT AGT AAT GTT CGC CAG ACC AAG AAG GCA CAG CCG AAG ACT CTA AAG CCT GTA
K   K   S   P   S   K   T   P   K   V   T   K   A   N   N   S   L   E   E   L   A   K   I   S   N   V   R   Q   T   K   K   A   Q   P   K   T   L   K   P   V
6241/2081                                  6271/2091                                  6301/2101                                  6331/2111
GAA GTA AAA CCT TTA AAG CCA GCC AGT CCA GTT CCA GAT GAT TTT GAG ATT CTT GAT CTG CTG GAA AAA GCC ACG GAG GCC AAA GAA CTG GAA CGC TCG AAG ACC AAG AGT CCA GCC GTA
E   V   K   P   L   K   P   A   S   P   V   P   D   D   F   E   I   L   D   L   L   E   K   A   T   E   A   K   E   L   E   R   S   K   T   K   S   P   A   V
6361/2121                                  6391/2131                                  6421/2141                                  6451/2151
GAA TCT ATC AGT CAG ACG CCC AAA GAA GCT ATA GTA GAA ATT TCC TTA CCA GTG GAA GAC ATT AAA AAT CTA CCC AAA ACC GGG TGT GAT AAG TCC TCA AAC AGC TCT CGC CGT GGT TCC
E   S   I   S   Q   T   P   K   E   A   I   V   E   I   S   L   P   V   E   D   I   K   N   L   P   K   T   G   C   D   K   S   S   N   S   S   R   R   G   S
6481/2161                                  6511/2171                                  6541/2181                                  6571/2191
CAA TCT AGT TTG ATC ATG TCG CGA CGA CAC TCG GAG ATT TCT CTA AAC GAG AAA CTC AAC CAG GAC GCA CTC GCA GCC TTG AAC CAA ATA GAA AAA GAA AGA GAG GCG GAG CAA GTG GAC
Q   S   S   L   I   M   S   R   R   H   S   E   I   S   L   N   E   K   L   N   Q   D   A   L   A   A   L   N   Q   I   E   K   E   R   E   A   E   Q   V   D
6601/2201                                  6631/2211                                  6661/2221                                  6691/2231
GAA CTT TTC CAG AGC ATG GTG GAG GAA ATG GAG CAG GAA CCG CAG CCC ACA GCT ATT GTC GAA CCT CCA GAA GAA GAC ATC GAT GCA GAT TCC TTG TGC ACA ACA ATT AGC AAG AGT CCC
```

FIGURE 14-3

DMICALlong.txt

```
E   L   F   Q   S   M   V   E   E   M   E   Q   E   P   Q   P   T   A   I   V   E   P   P   E   E   D   I   D   A   D   S   L   C   T   T   I   S   K   S   P
6721/2241                               6751/2251                               6781/2261                               6811/2271
AGT GCA CAG CCC GTT ACG GTG GTC AAA CGT GGC AGT TCC GAG GAT CAG AGC ATA GAG AAG CTT TTC AGC CAC TTT TCC GAC GAG ATG CTG GTG AAT GTG GAG TTT GAC TCT AAC GAT GAG
S   A   Q   P   V   T   V   V   K   R   G   S   S   E   D   Q   S   I   E   K   L   F   S   H   F   S   D   E   M   L   V   N   V   E   F   D   S   N   D   E
6841/2281                               6871/2291                               6901/2301                               6931/2311
CTT GTA GGA ATT ACA CCA AGG GCA ACA CTC GTT TCC CGA AAC ACA GAA GAT CGG GAC TAT CTG GAT AAA CTA GAG TCT TTG GAG CGC GAT GAG GAG ACT TTC CAA CCG GTT GTT GGG GAA
L   V   G   I   T   P   R   A   T   L   V   S   R   N   T   E   D   R   D   Y   L   D   K   L   E   S   L   E   R   D   E   E   T   F   Q   P   V   V   G   E
6961/2321                               6991/2331                               7021/2341                               7051/2351
AAA TTC ATA CAG GAA AAT GTC CAG GAC GAA GTG GAC GGA TTG CAT TTC CCA TCA CGC CCA CAA CGG AGA CCC AAA AGC AGT TCA TCT TCC AGT GAA CCA TCA CTT CCT GTG GCT CCT CAA
K   F   I   Q   E   N   V   Q   D   E   V   D   G   L   H   F   P   S   R   P   Q   R   R   P   K   S   S   S   S   S   S   E   P   S   L   P   V   A   P   Q
7081/2361                               7111/2371                               7141/2381                               7171/2391
AGG TTG GAA AAA AAG CTA TCG AAA TTA GAT CCG GAG GAT ATG CCT CCT TCT GTA CAG GAT CTA TTA CAA CAG GTT TAC CAG AAG AAC ATC CAA CCT GAA TTA GTG GAA GTA ATT CCA GTT
R   L   E   K   K   L   S   K   L   D   P   E   D   M   P   P   S   V   Q   D   L   L   Q   Q   V   Y   Q   K   N   I   Q   P   E   L   V   E   V   I   P   V
7201/2401                               7231/2411                               7261/2421                               7291/2431
GAA GGC AAA CAA ACT TTA AGG TTC CCC AGC ATG TTG GCG GAA GAA GAT GTA GAC GAA GTG GAT CAC TCT AAA GAA GGA ATC AAG AAA ATT GAA ACG GCT CCA GAA GAG GTT CGT AAG GTA
E   G   K   Q   T   L   R   F   P   S   M   L   A   E   E   D   V   D   E   V   D   H   S   K   E   G   I   K   K   I   E   T   A   P   E   E   V   R   K   V
7321/2441                               7351/2451                               7381/2461                               7411/2471
ACC GAG CCG GAG GAT GTT GCT CGG GTA ATT CCA AGT CCA ATC AAA CCG TCA ATA AGT CAG AGC AAC TCC CTC AAG AGC GAA AAT TCC TCT GGC AGC AGT TTA GTG GAA ATA CCA AAA ATA
T   E   P   E   D   V   A   R   V   I   P   S   P   I   K   P   S   I   S   Q   S   N   S   L   K   S   E   N   S   S   G   S   S   L   V   E   I   P   K   I
7441/2481                               7471/2491                               7501/2501                               7531/2511
ATT GCA CCA CCC AAG TCG AGT AGC AAG GAA AAC TCT TCG GAT TGG GAC AGG GAG AAG CTG CCC GCA TCA CCC ATG CCA CGA CGA AGA CTT CTG CCA AAT CAA ACG CCA TAT AAA GCC CCA
I   A   P   P   K   S   S   S   K   E   N   S   S   D   W   D   R   E   K   L   P   A   S   P   M   P   R   R   R   L   L   P   N   Q   T   P   Y   K   A   P
7561/2521                               7591/2531                               7621/2541                               7651/2551
AGC GTG GCT AGT AAG GAG AGC TCC TTA GAG TGG GAC ATG GAG AAG CTA CCC AAC AGT CCC ATG TTG CCA AGG AGG AAC AAG ATG CGA GCA ATC TCA CCT AGT ACC AAC CCT GTG CAA CTT
S   V   A   S   K   E   S   S   L   E   W   D   M   E   K   L   P   N   S   P   M   L   P   R   R   N   K   M   R   A   I   S   P   S   T   N   P   V   Q   L
7681/2561                               7711/2571                               7741/2581                               7771/2591
CTC AAC AAT TTA CCC TCC GAT GTA GAT GAC GAG GCG GCA CAG AGA CGC CTT ATC GAG GAT TTC GAG CAG GAG AGA CGT CAG GCT CTG ATC AAG CGG GAT GAG AAC TTC GAG GCC ATT GCG
L   N   N   L   P   S   D   V   D   D   E   A   A   Q   R   R   L   I   E   D   F   E   Q   E   R   R   Q   A   L   I   K   R   D   E   N   F   E   A   I   A
7801/2601                               7831/2611                               7861/2621                               7891/2631
GCG GAG CAA CGC AGG CGC GAC TCC TTA CAG AGC AGC AGT AAC TCG AGC AGC AAA CGC AGT TTA CCA CCG CCC ACG CCG CCC ATG ATG GCA TCG CGA CGC GGT ACA ACT CAA GAC ACA AAT
A   E   Q   R   R   R   D   S   L   Q   S   S   S   N   S   S   S   K   R   S   L   P   P   P   T   P   P   M   M   A   S   R   R   G   T   T   Q   D   T   N
7921/2641                               7951/2651                               7981/2661                               8011/2671
CGC ACC CAG GAT ACC GCA TCG CGG CAC GAG GGA ACG CCG CCC ATG TTC AAG AAG CTG GAT GTC GAT GGC AGC GGT ACA TCA ATG GAT TCC ACC TCC TGC TCC ACG CGG CGC AGT TCG TTT
R   T   Q   D   T   A   S   R   H   E   G   T   P   P   M   F   K   K   L   D   V   D   G   S   G   T   S   M   D   S   T   S   C   S   T   R   R   S   S   F
8041/2681                               8071/2691                               8101/2701                               8131/2711
GCA TTT ATA GAG TTA CAG GAT AAC AAA CCA GTG ATT GTG CCC ATG CCC AAG AAA CTG AAG TTG CCA AAG CCG GAG CCA CCT AGG TTC GTA CCC GAG CCA GTG GCC ACT GAT GAG CCT GTG
A   F   I   E   L   Q   D   N   K   P   V   I   V   P   M   P   K   K   L   K   L   P   K   P   E   P   P   R   F   V   P   E   P   V   A   T   D   E   P   V
8161/2721                               8191/2731                               8221/2741                               8251/2751
CCC GAG GTT TTT CAG GGT CGC GCT TGG CCC AAG ACA CAG TTG GAA GGA GAG GTC GAT CTA GGC GAT TCG GAT AAT GAA GAT GAA ACA GAA AAG CTA AAG AAA CAG CTG CCC GAA TAC GCT
P   E   V   F   Q   G   R   A   W   P   K   T   Q   L   E   G   E   V   D   L   G   D   S   D   N   E   D   E   T   E   K   L   K   K   Q   L   P   E   Y   A
8281/2761                               8311/2771                               8341/2781                               8371/2791
CGT TCG GAC TCT CCT CCT TCG GCT GCA TTC AAG AAT CGC AAG TGG CCG GAT GGG AAA ACA GTT TTT GAT AAA CGA GCC GAA TCT CTT GAG GAG GAA GAT ATC TTC GAA GGA TTA CCG TCC
R   S   D   S   P   P   S   A   A   F   K   N   R   K   W   P   D   G   K   T   V   F   D   K   R   A   E   S   L   E   E   E   D   I   F   E   G   L   P   S
8401/2801                               8431/2811                               8461/2821                               8491/2831
CCA AGG AAA AGA GGT TCT CAA AGA TTC ATG GAC AAG CCG CGC TCT CAA TCG CCA CAG CCT TTC AAA CCG CTG GCC AAC AGC TCC AGG AAA AGC TCA AAG TCT TTT AGC GAC CTT AAG AAA
P   R   K   R   G   S   Q   R   F   M   D   K   P   R   S   Q   S   P   Q   P   F   K   P   L   A   N   S   S   R   K   S   S   K   S   F   S   D   L   K   K
8521/2841                               8551/2851                               8581/2861                               8611/2871
GGA CCC TCC TTG CAA TCT CTG TCG GCG CAA TCC AGC CAG GAC ACG GAC ACA CTG TCC ACC ACC ACA ACA GTA GCC ACA GCT CGT CCT GCT AGC TAT GCC AAC TAC GAG GAC CCC ATG GAT
G   P   S   L   Q   S   L   S   A   Q   S   S   Q   D   T   D   T   L   S   T   T   T   T   V   A   T   A   R   P   A   S   Y   A   N   Y   E   D   P   M   D
8641/2881                               8671/2891                               8701/2901                               8731/2911
GCC AGT ACC CAA GCT TTG CTG GAT CGA AGC AAA CGG TTA CAC AAC CGC AAA AGA GAT TTC GTA AAC GAG CGA GTA GTG GAG CGC AAC CCC TAT ATG AGG GAT GTG CTT AGG AGC ACG GAT
A   S   T   Q   A   L   L   D   R   S   K   R   L   H   N   R   K   R   D   F   V   N   E   R   V   V   E   R   N   P   Y   M   R   D   V   L   R   S   T   D
8761/2921                               8791/2931                               8821/2941                               8851/2951
CGC CGT GAT TAC GAC GAC GTG GAT GAA GAT CTG ACT AGC TAC AGG CCA AGA CAT TAT GCC AGC TCC ACG CTA AAT CGT TTC CCC AAC ACC ACA ATA AGA AAA AGC AAC AAC TAC GAT TAC
R   R   D   Y   D   D   V   D   E   D   L   T   S   Y   R   P   R   H   Y   A   S   S   T   L   N   R   F   P   N   T   T   I   R   K   S   N   N   Y   D   Y
8881/2961                               8911/2971                               8941/2981                               8971/2991
```

FIGURE 14-4

DMICALlong.txt

```
CTC AGT CCC AGC AGT GAT TAT TTG AGC AGG AGA AGC TAC ATA CCG AGC GCC AGT GCA ACC AGC AGT TAC TAT CCA TCG ACC ACG CGT AGC TCC CAT TTG AGT GAC CTG TTC CGA CGA CGC
L   S   P   S   S   D   Y   L   S   R   R   S   Y   I   P   S   A   S   A   T   S   S   Y   Y   P   S   T   T   R   S   S   H   L   S   D   L   F   R   R   R
9001/3001                               9031/3011                               9061/3021                               9091/3031
AGC CCC GCC AGC GGA ACT GTA TCC GCA CTT TCC GGC TAC GGC AAC AAA GAG TCG TGC GTT ATC TCA ATC GGG CTG GCC TTA GAT CGA GTT GGC CAC TTG ATT GAA AGT AAA TGC ACC TGG
S   P   A   S   G   T   V   S   A   L   S   G   Y   G   N   K   E   S   C   V   I   S   I   G   L   A   L   D   R   V   G   H   L   I   E   S   K   C   T   W
9121/3041                               9151/3051                               9181/3061                               9211/3071
GTA CGA TCT ACG AAG GTT CAA ACC GAG TCC GAG AGC ACT TCA CCC GAC GAA GTG GAG CTC AAT TCT GCC ACT GAG ATA TCC ACC GAC TCT GAG TTT GAC AAC GAT GAG ATT ATA CGC CAG
V   R   S   T   K   V   Q   T   E   S   E   S   T   S   P   D   E   V   E   L   N   S   A   T   E   I   S   T   D   S   E   F   D   N   D   E   I   I   R   Q
9241/3081                               9271/3091                               9301/3101                               9331/3111
GCG CCC AAA ATC TTC ATC GAT GAC ACC CAT CTA AGG AAG CCC ACC AAG GTT CAG ATC AAG TCC ACC ATG ATC GGA CCC AAT GCA GCT TCC GCC GGA CTC CAT CAG AAG CAG TTG GCG GCG
A   P   K   I   F   I   D   D   T   H   L   R   K   P   T   K   V   Q   I   K   S   T   M   I   G   P   N   A   A   S   A   G   L   H   Q   K   Q   L   A   A
9361/3121                               9391/3131                               9421/3141                               9451/3151
CGT GAG AAG GGC GGC AGC TAC CTC CAG AAG TAC CAA CCA CAA CCG CCA CTG CCA CAG TTT AGA CCG TTG GTC CAG GTG GAT CCC ACC CTG CTC ATT GGC AGC CAG CGC GCT CCT CTT CAG
R   E   K   G   G   S   Y   L   Q   K   Y   Q   P   Q   P   P   L   P   Q   F   R   P   L   V   Q   V   D   P   T   L   L   I   G   S   Q   R   A   P   L   Q
9481/3161                               9511/3171                               9541/3181                               9571/3191
AAT CCA CGG CCA GGA GAC TAC TTG CTA AAC AAG ACG GCC AGT ACG GAG GGT ATC GCC TCA AAA AAG AGC CTG GGG CTA AAA AAG CGC TAT CTG CTG GGT GAG CCG GCC AAT GGC AAT AAG
N   P   R   P   G   D   Y   L   L   N   K   T   A   S   T   E   G   I   A   S   K   K   S   L   G   L   K   K   R   Y   L   L   G   E   P   A   N   G   N   K
9601/3201                               9631/3211                               9661/3221                               9691/3231
ATC CAG AAG TCC GGA TCC ACT TCA GTG CTG GAT TCA CGC ATT CGC AGC TTC CAG TCG AAC ATA TCG GAG TGC CAG AAG CTT TTG AAT CCC AGC AGC GAC ATA AGT GCC GGC ATG CGA ACC
I   Q   K   S   G   S   T   S   V   L   D   S   R   I   R   S   F   Q   S   N   I   S   E   C   Q   K   L   L   N   P   S   S   D   I   S   A   G   M   R   T
9721/3241                               9751/3251                               9781/3261                               9811/3271
TTC CTC GAT CGC ACA AAG TTG GGC GAA GGC AGC CAG ACG ACA CCC GGA CAG ACG AAC GAA CTA ATC CGT TCC GCC ACC AGC AAT GTG ATT AAC GAT CTG CGC GTG GAG CTT CGG ATA CAG
F   L   D   R   T   K   L   G   E   G   S   Q   T   T   P   G   Q   T   N   E   L   I   R   S   A   T   S   N   V   I   N   D   L   R   V   E   L   R   I   Q
9841/3281                               9871/3291                               9901/3301                               9931/3311
AAA ACT GGC TCC AGC CAC TCC ACG GAC AAC GAG AAG GAA AAC GTT TTC GTG AAC TGT AAG AAC GAG CTG AAC AAG GGG ATG GAA TAC ACG GAT GCG GTC AAT GCC ACG CTG CTG GAC CAG
K   T   G   S   S   H   S   T   D   N   E   K   E   N   V   F   V   N   C   K   N   E   L   N   K   G   M   E   Y   T   D   A   V   N   A   T   L   L   D   Q
9961/3321                               9991/3331                               10021/3341                              10051/3351
CTG GCC AGA AAA AGT TCA CCC ACC ACG CCG ACG AAT AAG ACG GTG GTC GAG GTT ATT GAC CTG GTT ACA CCT GAG AAG CCA ATT GAC ATT ATC GAT CTA ACG GCA CTG GAA ACG CCG AAA
L   A   R   K   S   S   P   T   T   P   T   N   K   T   V   V   E   V   I   D   L   V   T   P   E   K   P   I   D   I   I   D   L   T   A   L   E   T   P   K
10081/3361                              10111/3371                              10141/3381                              10171/3391
AAG CAG TTG GTC GAT GGT AGC GCC ATG GAT GTA GAT GAA CGC CTC ACA CCC GAT AGC AAC AAA ATC AGC GAA CTG CAG CAG GAA GTG AAG GAG GAA CCC AAG CCG GAT GTC TCT AGG GAT
K   Q   L   V   D   G   S   A   M   D   V   D   E   R   L   T   P   D   S   N   K   I   S   E   L   Q   Q   E   V   K   E   E   P   K   P   D   V   S   R   D
10201/3401                              10231/3411                              10261/3421                              10291/3431
GTG AAA GAA TGC ATA CCA GAT ATA CTG GGA CAC ATT AAG GAG GGA ACG GGA TCG AAG GAG CCA GGT GGA GAG GAC CAA CAG AGC CTG CTG GAG CAG TCG GAC GAA GAG AAG CGC GAC TCA
V   K   E   C   I   P   D   I   L   G   H   I   K   E   G   T   G   S   K   E   P   G   G   E   D   Q   Q   S   L   L   E   Q   S   D   E   E   K   R   D   S
10321/3441                              10351/3451                              10381/3461                              10411/3471
CCG GAA AAG GAT GTG GCC GAA CAT GAG CTT TAT GAA CCG GAC AGT GTG CAG ATC CAG GTG CCC AAT ATC CCA TGG GAA AAA AGC AAG CCG GAG GTC ATG TCT ACC ACC GGC AGC AGT GGC
P   E   K   D   V   A   E   H   E   L   Y   E   P   D   S   V   Q   I   Q   V   P   N   I   P   W   E   K   S   K   P   E   V   M   S   T   T   G   S   S   G
10441/3481                              10471/3491                              10501/3501                              10531/3511
TCC ATC TGC TCA AGC TCA GAC TCT TCT AGT ATT GAA GAC ATC CAG CAC TAC ATT TTG GAG TCC ACA ACT AGT CCA GAT ACT CAG ACA GTT GGC GGA AAG CAC AAT GTG CCC CGT TTG GAG
S   I   C   S   S   S   D   S   S   S   I   E   D   I   Q   H   Y   I   L   E   S   T   T   S   P   D   T   Q   T   V   G   G   K   H   N   V   P   R   L   E
10561/3521                              10591/3531                              10621/3541                              10651/3551
GTG CAC GAC ACA AGT GGT GCC CTG ATG CAG GTG GAC AGC CTG ATG ATT GTG AAC GGA AAG TAT ATT GGG GAT CCC GAG GAT GTC AAG TTC TTG GAT ATG CCG GCC AAT GTT ATT GTT CCG
V   H   D   T   S   G   A   L   M   Q   V   D   S   L   M   I   V   N   G   K   Y   I   G   D   P   E   D   V   K   F   L   D   M   P   A   N   V   I   V   P
10681/3561                              10711/3571                              10741/3581                              10771/3591
CCA GCA CCG GCG CTT AAA ACG AAT GAG CTG GAT ATG GAG GAT GAC CAA GAG GCG GAG GCG GAA CCA GTA ACT GCT ACT CCG GAG CCG GTG GAA TGT ACG GTC ATC GAG GCT GAG CGC CGT
P   A   P   A   L   K   T   N   E   L   D   M   E   D   D   Q   E   A   E   A   E   P   V   T   A   T   P   E   P   V   E   C   T   V   I   E   A   E   R   R
10801/3601                              10831/3611                              10861/3621                              10891/3631
GTT ACT GCT CCC CCT CCT TTG CCA GAG ATG GGT CCA CCC AAA CTG AAG TTC GAT AGC AAA AAT GAG AAC AAG ATC GAG AGC TTG AAG AAT CTT CCG TTG ATC GTA GAG AGC AAT GTG GAG
V   T   A   P   P   P   L   P   E   M   G   P   P   K   L   K   F   D   S   K   N   E   N   K   I   E   S   L   K   N   L   P   L   I   V   E   S   N   V   E
10921/3641                              10951/3651                              10981/3661                              11011/3671
CAC AGT CAG GCA GTG AAA CCC ATT ACT CTT AAC TTA AGC AAT CTG GCC AGG ACG CCG GAT ACA CCA ACC ACG CCC ACG GCG CAC GAT AGC GAT AAA ACA CCC ACT GGG GAA ATT CTC TCG
H   S   Q   A   V   K   P   I   T   L   N   L   S   N   L   A   R   T   P   D   T   P   T   T   P   T   A   H   D   S   D   K   T   P   T   G   E   I   L   S
11041/3681                              11071/3691                              11101/3701                              11131/3711
CGA GGA TCT GAC TCA GAA ACC GAG CAC ACT GGC ACT GGT CAG GTA CTA ACG GAG ACG GAA CTC TCC GAC TGG ACG GCC GAC GAC TGT ATC TCG GAG AAC TTT GTT GAC TTG GAG TTC GCG
R   G   S   D   S   E   T   E   H   T   G   T   G   Q   V   L   T   E   T   E   L   S   D   W   T   A   D   D   C   I   S   E   N   F   V   D   L   E   F   A
```

FIGURE 14-5

DMICALlong.txt

```
11161/3721                     11191/3731                     11221/3741                     11251/3751
CTT AAC TCT AAC AAG GGT ACG ATA AAA CGG CGC AAG GAT CGA CGA CGC AGT GGA GCA AGC AAA CTT CCC AGT GGC AAC GAG GTA ATC CAC GAG CTG GCC AGG CAG GCG CCA GTG GTG CAA
L   N   S   N   K   G   T   I   K   R   R   K   D   R   R   R   S   G   A   S   K   L   P   S   G   N   E   V   I   H   E   L   A   R   Q   A   P   V   V   Q
11281/3761                     11311/3771                     11341/3781                     11371/3791
ATG GAT GGA ATT CTT AGT GCC ATC GAC ATT GAT GAC ATA GAG TTC ATG GAC ACG GGT TCG GAG GGT TCT TGT GCT GAA GCT TAT CCC GCA ACA AAT ACA GCT CTC ATT CAG AAT AGA GGT
M   D   G   I   L   S   A   I   D   I   D   D   I   E   F   M   D   T   G   S   E   G   S   C   A   E   A   Y   P   A   T   N   T   A   L   I   Q   N   R   G
11401/3801                     11431/3811                     11461/3821                     11491/3831
TAC ATG GAG TAC ATC GAG GCG GAG CCG AAA AAG ACG ACC CGC AAG GCA GCT CCA CCA TCC AGT TAC CCA GGA AAT TTA CCG CCT TTA ATG ACG AAG CGG GAC GAG AAA CTG GGC GTT GAT
Y   M   E   Y   I   E   A   E   P   K   K   T   T   R   K   A   A   P   P   S   S   Y   P   G   N   L   P   P   L   M   T   K   R   D   E   K   L   G   V   D
11521/3841                     11551/3851                     11581/3861                     11611/3871
TAC ATT GAG CAG GGG GCG TAC ATA ATG CAC GAT GAT GCA AAG ACG CCT GTG AAT GAG GTG GCT CCT GCC ATG ACC CAG TCG CTA ACT GAC TCA ATC ACG CTC AAT GAA CTG GAT GAT GAC
Y   I   E   Q   G   A   Y   I   M   H   D   D   A   K   T   P   V   N   E   V   A   P   A   M   T   Q   S   L   T   D   S   I   T   L   N   E   L   D   D   D
11641/3881                     11671/3891                     11701/3901                     11731/3911
AGT ATG ATA ATA TCC CAA ACC CAG CCA ACG ACA ACG GAG GAA AGT GAG GCA CTG ACG GTG GTC ACC AGT CCA CTT GAC ACG TCC TCG CCC AGG GTT CTC GAT CAA TTT GCA TCC ATG TTG
S   M   I   I   S   Q   T   Q   P   T   T   T   E   E   S   E   A   L   T   V   V   T   S   P   L   D   T   S   S   P   R   V   L   D   Q   F   A   S   M   L
11761/3921                     11791/3931                     11821/3941                     11851/3951
GCG GCG GGA AAA GGT GAC TCC ACA CCC AGT AGC TCA GAG CAA CAA CCA AAG ACG TCT ACG GTG ACG AGC AGC AGC ACT GGG CCC AAC TCC TCG ACA ACA GGA AAC GTC TCG AAG GAG CCG
A   A   G   K   G   D   S   T   P   S   S   S   E   Q   Q   P   K   T   S   T   V   T   S   S   S   T   G   P   N   S   S   T   T   G   N   V   S   K   E   P
11881/3961                     11911/3971                     11941/3981                     11971/3991
CAG GAG GAG GAC CTG CAA ATC CAG TTT GAG TAT GTT CGA GCA CTG CAG CAG CGG ATA TCG CAG ATC AGC ACC CAA CGG CGT AAG AGC TCT AAG GGA GAG GCA CCT AAC CTG CAG CTA AAC
Q   E   E   D   L   Q   I   Q   F   E   Y   V   R   A   L   Q   Q   R   I   S   Q   I   S   T   Q   R   R   K   S   S   K   G   E   A   P   N   L   Q   L   N
12001/4001                     12031/4011                     12061/4021                     12091/4031
AGT AGC GCA CCT GTG ATA GAA TCA GCC GAG GAT CCG GCC AAG CCC GCA GAG GAG CCT CTG GTC TCA ATG CGA CCG CGG ACC ACC AGC ATT TCC GGA AAG GTA CCG GAG ATA CCC ACA CTT
S   S   A   P   V   I   E   S   A   E   D   P   A   K   P   A   E   E   P   L   V   S   M   R   P   R   T   T   S   I   S   G   K   V   P   E   I   P   T   L
12121/4041                     12151/4051                     12181/4061                     12211/4071
AGC AGC AAG CTG GAA GAG ATA ACC AAA GAA CGC ACT AAG CAA AAG GAT CTG ATT CAC GAC CTA GTC ATG GAC AAG TTG CAG TCG AAG AAG CAG CTA AAC GCT GAG AAG CGT CTG CAC CGG
S   S   K   L   E   E   I   T   K   E   R   T   K   Q   K   D   L   I   H   D   L   V   M   D   K   L   Q   S   K   K   Q   L   N   A   E   K   R   L   H   R
12241/4081                     12271/4091                     12301/4101                     12331/4111
AGT CGA CAG CGC AGT TTG CTG ACC AGT GGC TAT GCC AGT GGC TCC AGC CTT AGT CCG ACG CCC AAG CTG GCT GCT GCT TGC AGT CCG CAG GAT TCC AAC TGC TCT AGC CAA GCG CAC TAC
S   R   Q   R   S   L   L   T   S   G   Y   A   S   G   S   S   L   S   P   T   P   K   L   A   A   A   C   S   P   Q   D   S   N   C   S   S   Q   A   H   Y
12361/4121                     12391/4131                     12421/4141                     12451/4151
CAC GCC TCC ACG GCG GAG GAG GCC CCG AAG CCG CCG GCG GAA AGG CCG TTG CAG AAG TCC GCC ACG TCC ACC TAT GTG TCG CCT TAT CGC ACT GTC CAA GCG CCC ACA CGT AGT GCT GAT
H   A   S   T   A   E   E   A   P   K   P   P   A   E   R   P   L   Q   K   S   A   T   S   T   Y   V   S   P   Y   R   T   V   Q   A   P   T   R   S   A   D
12481/4161                     12511/4171                     12541/4181                     12571/4191
CTC TAT AAG CCG CGC CCC TTC AGC GAA CAC ATC GAT TCG AAC GCT CTG GCG GGT TAC AAG CTC GGC AAG ACG GCC TCG TTT AAT GGC GGC AAG TTG GGC GAC TTT GCG AAA CCC ATT GCC
L   Y   K   P   R   P   F   S   E   H   I   D   S   N   A   L   A   G   Y   K   L   G   K   T   A   S   F   N   G   G   K   L   G   D   F   A   K   P   I   A
12601/4201                     12631/4211                     12661/4221                     12691/4231
CCG GCG AGA GTT AAC CGA GGA GGA GGT GTC GCG ACC GCG GAT ATA GCC AAT ATT TCC GCG TCG ACG GAG AAC CTA AGA AGC GAG GCC AGG GCC AGG GCT CGT CTT AAG TCT AAC ACA GAG
P   A   R   V   N   R   G   G   G   V   A   T   A   D   I   A   N   I   S   A   S   T   E   N   L   R   S   E   A   R   A   R   A   R   L   K   S   N   T   E
12721/4241                     12751/4251                     12781/4261                     12811/4271
CTG GGC CTT AGT CCC GAG GAA AAG ATG CAG CTA ATA CGT TCA AGA TTG CAC TAC GAC CAA AAC AGA TCT CTG AAG CCG AAG CAA CTG GAG GAG ATG CCA TCC GGG GAT CTG GCG GCA CGT
L   G   L   S   P   E   E   K   M   Q   L   I   R   S   R   L   H   Y   D   Q   N   R   S   L   K   P   K   Q   L   E   E   M   P   S   G   D   L   A   A   R
12841/4281                     12871/4291                     12901/4301                     12931/4311
GCC CGC AAA ATG AGT GCC TCG AAG AGC GTC AAT GAT CTG GCC TAC ATG GTG GGA CAG CAG CAG CAG CAG CAG GTT GAG AAG GAT GCC GTG CTC CAA GCC AAG GCG GCT GAC TTT ACA TCC
A   R   K   M   S   A   S   K   S   V   N   D   L   A   Y   M   V   G   Q   Q   Q   Q   Q   Q   V   E   K   D   A   V   L   Q   A   K   A   A   D   F   T   S
12961/4321                     12991/4331                     13021/4341                     13051/4351
GAT CCC AAT TTG GCG TCC GGT GGT CAG GAG AAG GCA GGC AAA ACT AAG TCC GGA CGC AGG CCA AAG GAT CCG GAG CGG CGT AAG AGT CTC ATA CAG TCG TTG TCC AGC TTC TTC CAA AAG
D   P   N   L   A   S   G   G   Q   E   K   A   G   K   T   K   S   G   R   R   P   K   D   P   E   R   R   K   S   L   I   Q   S   L   S   S   F   F   Q   K
13081/4361                     13111/4371                     13141/4381                     13171/4391
GGA TCT GGA TCC GCG GCC TCC AGT TCC AAG GAG CAG GGC GGC GCT GTG GCT GCC GTC CAC TCT GAA CAG TCA GAG CGA CCA GGC ACC AGC AGC AGC GGC ACG CCC ACA ATA TCG GAT GCG
G   S   G   S   A   A   S   S   S   K   E   Q   G   G   A   V   A   A   V   H   S   E   Q   S   E   R   P   G   T   S   S   S   G   T   P   T   I   S   D   A
13201/4401                     13231/4411                     13261/4421                     13291/4431
GCG GGT GGA GGC GGA GGA GGA GGT GGC GTC TTC AGC AGA TTC CGC ATC TCG CCC AAG TCC AAG GAG AAG TCA AAG TCT TGC TTT GAT CTG AGG AAT TTC GGT TTT GGT GAC AAG GAT ATG
A   G   G   G   G   G   G   G   G   V   F   S   R   F   R   I   S   P   K   S   K   E   K   S   K   S   C   F   D   L   R   N   F   G   F   G   D   K   D   M
13321/4441                     13351/4451                     13381/4461                     13411/4471
CTG GTC TGC AAT GCA GCA TCT CCA GCA GGA GCC ACA TCC GCA TCA CAG AAA AAT CAC TCG CAA GAG TAT CTG AAC ACC ACG AAC AAC AGT CGC TAT CGA AAG CAA ACG AAC ACT GCG AAA
```

FIGURE 14-6

DMICALlong.txt

```
L   V   C   N   A   A   S   P   A   G   A   T   S   A   S   Q   K   N   H   S   Q   E   Y   L   N   T   T   N   N   S   R   Y   R   K   Q   T   N   T   A   K
13441/4481                                  13471/4491                                  13501/4501                                  13531/4511
CCG AAA CCC GAA TCG TTC TCT TCA TCC AGT CCG CAG CTC TAT ATA CAC AAG CCC CAC CAC CTG GCC GCA GCT CAT CCC AGT GCC CTG GAC GAC CAG ACA CCA CCA CCC ATA CCG CCT CTT
P   K   P   E   S   F   S   S   S   S   P   Q   L   Y   I   H   K   P   H   H   L   A   A   A   H   P   S   A   L   D   D   Q   T   P   P   P   I   P   P   L
13561/4521                                  13591/4531                                  13621/4541                                  13651/4551
CCA CTG AAT TAT CAG AGA TCC GAT GAT GAG AGC TAC GCT AAC GAG ACA CGA GAG CAT AAG AAG CAA CGT GCC ATA TCG AAG GCT TCA CGA CAA GCT GAG CTC AAG CGA TTG CGA ATC GCT
P   L   N   Y   Q   R   S   D   D   E   S   Y   A   N   E   T   R   E   H   K   K   Q   R   A   I   S   K   A   S   R   Q   A   E   L   K   R   L   R   I   A
13681/4561                                  13711/4571                                  13741/4581                                  13771/4591
CAA GAG ATT CAG CGG GAA CAG GAG GAG ATC GAG GTG CAA CTG AAG GAT CTG GAG GCA CGC GGC GTG CTT ATT GAG AAG GCC TTG CGA GGC GAG GCG CAG AAT ATT GAA AAC CTG GAT GCG
Q   E   I   Q   R   E   Q   E   E   I   E   V   Q   L   K   D   L   E   A   R   G   V   L   I   E   K   A   L   R   G   E   A   Q   N   I   E   N   L   D   A
13801/4601                                  13831/4611                                  13861/4621                                  13891/4631
ACA AAG GAC AAC GAC GAG AAG CTA CTT AAG GAA CTT TTG GAG ATT TGG CGC AAC ATC ACA GCA CTC AAG AAA CGC GAT GAG GAA CTG ACT ATA AGG CAA CAG GAA CTG CAA CTG GAG TAT
T   K   D   N   D   E   K   L   L   K   E   L   L   E   I   W   R   N   I   T   A   L   K   K   R   D   E   E   L   T   I   R   Q   Q   E   L   Q   L   E   Y
13921/4641                                  13951/4651                                  13981/4661                                  14011/4671
CGG CAT GCC CAG CTG AAG GAA GAG CTC AAT CTG CGC TTG TCC TGC AAC AAA CTG GAC AAA AGC TCT GCC GAT GTG GCC GCC GAG GGA GCA ATT CTC AAC GAG ATG CTG GAA ATT GTC GCC
R   H   A   Q   L   K   E   E   L   N   L   R   L   S   C   N   K   L   D   K   S   S   A   D   V   A   A   E   G   A   I   L   N   E   M   L   E   I   V   A
14041/4681                                  14071/4691                                  14101/4701                                  14131/4711
AAG CGA GCC GCC CTA CGA CCC ACA GCC TCC CAG CTC GAC CTC ACG GCA GCG GGA TCA GCA TCC ACG TCC GCC GAG GCA ACG GGC ATT AAG CTG ACG GGA CAA CCG CAT GAC CAC GAA GAA
K   R   A   A   L   R   P   T   A   S   Q   L   D   L   T   A   A   G   S   A   S   T   S   A   E   A   T   G   I   K   L   T   G   Q   P   H   D   H   E   E
14161/4721
TCG ATC ATT TGA
S   I   I   *
ATAGCCAACTTTCCCTGCAGATTGCTGGTCTGCGAAGTGGTGCATGGACTGTTGAGCCTAAATACTTAAGATGAAACAAGGGAACAGGGGAGAGCGTACCAAAAAGGCAACTGAATATACAGAATTAACAATACTAAAACCCGTGAGCTGTAACCGAAAC
GAGCAACTAACTATTCAAAGTGGATAAGCTGGCTAAGAACAATTTCAATATTTATATAAGTAGAGAGCCTAACTAATATGGGAGAAGCGCGCAAGGTTAATGTAAAAACAGGCAGAGATGAAATCAAAACCAAAAAACTAATTTCGCAATTTAAAAGTAG
GGATTACGAATGTTTTTGCGAATACCTTTAGTTTCGTATTCGACCTAATTGTACGTTTTGTTAGTCGTTGCCACTGAAATCTGTCTGATTTTTGTATGTCCTATGCAATATGTTGTGTTTTTATTTAATGTAATGTAGTGAGCGTGTGCAAAGGGTATGT
GATTTTGAGCAAATGTAAATGGTGTATTTGAGTTTTCTTAGGATATTTTCACTTAGCGATTGTTTGTCGATTTCTATGTAAACCGTTTGAGTGTATTATTATTTATTTGACAAAGCACCTTCCGCACATCCTCCCGATCAGCACACCCATGTATGTATTA
TGTCATTTGCCCCCAATCGAATTCAAAGTGAACACAAAGAGATGCACACAAATATTTTCTAAATGCGAGAGCCTACACATAGCGATGTAAGCGAGAAAGCAACCATTAACTGTATACTAACGAAATCATTTAAATTATTGCGTAACTTTATTTATAACCT
GCATATATAAAAACATATTTATCGAATTTATTATGTAACTTTTCGATTTATTTTGAATTGAATAAAGATTTTATGAAACCTAA
```

FIGURE 14-7

```
DMICALmedium.txt
GCTCTGCATCTCGCAACGAATCGTTCCGAATAAAAAATTAAATAAATAAACACAAAAAATTGTGATATAAGCGATAAATATTGTGATAATTCGGACAAACGTGTCAATTGCATTGCATCAGA
1/1                                     31/11                                   61/21                                   91/31
ATG AGC CGC CAA CAC CAG CGG CAC CAC CAG CAG CAT CAC CAC CTG CCG CCG CAC CAG CAA CCG CAG CAG CAG ATG CCG CAA CAA CAG CAG CAG CTG ACG GCG CAG CAG CAG CAA CAA CAG
M   S   R   Q   H   Q   R   H   H   Q   Q   H   H   H   L   P   P   H   Q   Q   P   Q   Q   Q   M   P   Q   Q   Q   Q   Q   L   T   A   Q   Q   Q   Q   Q   Q
121/41                                  151/51                                  181/61                                  211/71
CAG CTG CTG ATG GCG GAG CAC GCG GCG GCC GCG GAG GCG GCG GAG CTA TTC GAC CTG CTG TGC GTG GCC ACA ACG ATG CGC CAG ATC CTG GCG CTC CAT CGG GCC ATG TGC GAG GCT GTG
Q   L   L   M   A   E   H   A   A   A   A   E   A   A   E   L   F   D   L   L   C   V   A   T   T   M   R   Q   I   L   A   L   H   R   A   M   C   E   A   V
241/81                                  271/91                                  301/101                                 331/111
GGA TTG AGA CCC TCG CCT CTG AAC GAC TTC TAC CCA CGG CTA AAG GCC AAG GTG CGT TCG TGG AAG GCG CAG GCC CTG TGG AAG AAG TTC GAC GCC AGA GCT GCC CAT AGA GTC TAC GGC
G   L   R   P   S   P   L   N   D   F   Y   P   R   L   K   A   K   V   R   S   W   K   A   Q   A   L   W   K   K   F   D   A   R   A   A   H   R   V   Y   G
361/121                                 391/131                                 421/141                                 451/151
AAG GGA GCT GCC TGT ACT GGC ACA CGC GTC CTG GTC ATC GGA GCA GGG CCC TGT GGA CTG CGC ACC GCC ATC GAG GCC CAA CTG CTG GGC GCC AAG GTG GTG GTG CTG GAG AAA CGC GAT
K   G   A   A   C   T   G   T   R   V   L   V   I   G   A   G   P   C   G   L   R   T   A   I   E   A   Q   L   L   G   A   K   V   V   V   L   E   K   R   D
481/161                                 511/171                                 541/181                                 571/191
CGC ATC ACC CGG AAC AAT GTG CTC CAT CTG TGG CCA TTC GTC ATC ACG GAT CTG CGC AAC TTG GGC GCA AAG AAG TTC TAC GGC AAG TTT TGC GCC GGC TCC ATC GAT CAC ATC TCC ATT
R   I   T   R   N   N   V   L   H   L   W   P   F   V   I   T   D   L   R   N   L   G   A   K   K   F   Y   G   K   F   C   A   G   S   I   D   H   I   S   I
601/201                                 631/211                                 661/221                                 691/231
CGG CAG CTG CAG TGC ATG CTG CTC AAG GTG GCG CTG CTC CTG GGC GTA GAG ATC CAC GAG GGA GTC AGT TTT GAT CAC GCT GTA GAG CCC TCT GGC GAT GGC GGC GGA TGG AGG GCA GCT
R   Q   L   Q   C   M   L   L   K   V   A   L   L   L   G   V   E   I   H   E   G   V   S   F   D   H   A   V   E   P   S   G   D   G   G   G   W   R   A   A
721/241                                 751/251                                 781/261                                 811/271
GTT ACT CCC GCA GAT CAT CCT GTA TCT CAC TAC GAA TTC GAT GTG TTG ATC GGA GCG GAT GGC AAG CGG AAT ATG CTG GAC TTT AGG AGG AAG GAG TTC CGC GGG AAG CTG GCC ATC GCT
V   T   P   A   D   H   P   V   S   H   Y   E   F   D   V   L   I   G   A   D   G   K   R   N   M   L   D   F   R   R   K   E   F   R   G   K   L   A   I   A
841/281                                 871/291                                 901/301                                 931/311
ATT ACA GCG AAC TTT ATC AAC AAG AAG ACG GAG GCG GAG GCT AAA GTA GAG GAG ATC AGT GGG GTG GCT TTC ATC TTC AAC CAG GCC TTC TTC AAG GAG CTG TAC GGG AAG ACG GGC ATC
I   T   A   N   F   I   N   K   K   T   E   A   E   A   K   V   E   E   I   S   G   V   A   F   I   F   N   Q   A   F   F   K   E   L   Y   G   K   T   G   I
961/321                                 991/331                                 1021/341                                1051/351
GAC CTG GAA AAC ATC GTC TAC TAC AAG GAC GAG ACG CAC TAC TTC GTG ATG ACG GCC AAG AAG CAC AGT CTA ATT GAC AAG GGC GTT ATT ATC GAG GAT ATG GCC GAT CCC GGC GAG CTT
D   L   E   N   I   V   Y   Y   K   D   E   T   H   Y   F   V   M   T   A   K   K   H   S   L   I   D   K   G   V   I   I   E   D   M   A   D   P   G   E   L
1081/361                                1111/371                                1141/381                                1171/391
CTC GCC CCA GCC AAT GTG GAT ACA CAA AAG CTG CAC GAC TAT GCA CGC GAG GCT GCG GAG TTC TCC ACC CAA TAC CAA ATG CCA AAC CTG GAG TTC GCT GTT AAT CAC TAC GGC AAA CCA
L   A   P   A   N   V   D   T   Q   K   L   H   D   Y   A   R   E   A   A   E   F   S   T   Q   Y   Q   M   P   N   L   E   F   A   V   N   H   Y   G   K   P
1201/401                                1231/411                                1261/421                                1291/431
GAT GTG GCC ATG TTC GAC TTC ACA TCG ATG TTT GCC GCC GAG ATG TCC TGT CGG GTG ATT GTG CGC AAA GGA GCT CGC CTG ATG CAG TGC CTC GTG GGT GAC AGT CTG CTC GAG CCG TTT
D   V   A   M   F   D   F   T   S   M   F   A   A   E   M   S   C   R   V   I   V   R   K   G   A   R   L   M   Q   C   L   V   G   D   S   L   L   E   P   F
1321/441                                1351/451                                1381/461                                1411/471
TGG CCC ACT GGA TCG GGT TGT GCC CGT GGA TTC TTA TCC AGC ATG GAT GCT GCC TAT GCC ATC AAG CTT TGG TCC AAC CCG CAG AAC AGC ACA CTT GGC GTT CTG GCG CAG CGC GAA AGC
W   P   T   G   S   G   C   A   R   G   F   L   S   S   M   D   A   A   Y   A   I   K   L   W   S   N   P   Q   N   S   T   L   G   V   L   A   Q   R   E   S
1441/481                                1471/491                                1501/501                                1531/511
ATC TAC CGG CTG CTT AAC CAG ACC ACG CCG GAC ACC CTG CAG CGG GAC ATC AGT GCC TAT ACC GTG GAT CCG GCC ACG CGC TAT CCG AAT CTG AAC AGG GAG TCG GTC AAT AGC TGG CAG
I   Y   R   L   L   N   Q   T   T   P   D   T   L   Q   R   D   I   S   A   Y   T   V   D   P   A   T   R   Y   P   N   L   N   R   E   S   V   N   S   W   Q
1561/521                                1591/531                                1621/541                                1651/551
GTC AAA CAT CTG GTC GAC ACG GAC GAC CCG TCC ATT CTG GAG CAG ACC TTC ATG GAC ACG CAT GCT CTG CAG ACC CCG CAT TTG GAC ACA CCG GGC AGA CGC AAG CGA CGC AGT GGA GAC
V   K   H   L   V   D   T   D   D   P   S   I   L   E   Q   T   F   M   D   T   H   A   L   Q   T   P   H   L   D   T   P   G   R   R   K   R   R   S   G   D
1681/561                                1711/571                                1741/581                                1771/591
TTG CTG CCC CAG GGT GCC ACG TTG CTG AGA TGG ATA AGT GCC CAG CTG CAT TCC TAT CAG TTT ATT CCC GAA CTC AAG GAG GCT TCG GAT GTG TTC CGG AAT GGA CGC GTT CTG TGT GCG
L   L   P   Q   G   A   T   L   L   R   W   I   S   A   Q   L   H   S   Y   Q   F   I   P   E   L   K   E   A   S   D   V   F   R   N   G   R   V   L   C   A
1801/601                                1831/611                                1861/621                                1891/631
CTT ATC AAT CGC TAT CGT CCT GAT CTC ATC GAC TAC GCT GCC ACC AAG GAC ATG AGT CCC GTG GAG TGC AAT GAG CTG TCA TTC GCC GTC CTA GAG CGC GAA CTC CAC ATC GAT CGC GTC
L   I   N   R   Y   R   P   D   L   I   D   Y   A   A   T   K   D   M   S   P   V   E   C   N   E   L   S   F   A   V   L   E   R   E   L   H   I   D   R   V
1921/641                                1951/651                                1981/661                                2011/671
ATG AGT GCC AAA CAG TCG CTG GAC TTG ACC GAG CTG GAG TCG CGA ATC TGG CTC AAC TAT TTG GAC CAG ATC TGC GAC TTG TTT CGC GGC GAG ATC CCC CAT ATC AAG CAC CCC AAG ATG
M   S   A   K   Q   S   L   D   L   T   E   L   E   S   R   I   W   L   N   Y   L   D   Q   I   C   D   L   F   R   G   E   I   P   H   I   K   H   P   K   M
2041/681                                2071/691                                2101/701                                2131/711
GAC TTT AGC GAT TTG CGC CAG AAG TAT CGT ATC AAC CAT ACG CAT GCC CAA CCC GAC TTC TCC AAG CTG CTG GCA ACG AAA CCC AAG GCC AAG TCG CCG ATG CAG GAT GCT GTG GAC ATA
D   F   S   D   L   R   Q   K   Y   R   I   N   H   T   H   A   Q   P   D   F   S   K   L   L   A   T   K   P   K   A   K   S   P   M   Q   D   A   V   D   I
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 15-1

```
DMICALmedium.txt
CCC ACG ACA GTG CAG CGG CGC TCG GTG CTC GAG GAG GAG CGA GCC AAG CGG CAG CGT CGC CAC GAG CAG CTT CTT AAC ATC GGT GGA GGG GCA GCA GGA GCC GCC GCC GGA GTT GCC GGA
P   T   T   V   Q   R   R   S   V   L   E   E   E   R   A   K   R   Q   R   R   H   E   Q   L   L   N   I   G   G   G   A   A   G   A   A   A   G   V   A   G
2281/761                                2311/771                                2341/781                                2371/791
AGC GGG ACA GGA ACC ACA ACG CAG GGT CAA AAC GAT ACG CCA CGC CGG TCC AAG AAG CGC CGT CAG GTT GAC AAA ACC GCC AAT ATT GAG GAG CGC CAG CAG CGC TTG CAG GAG ATC GAG
S   G   T   G   T   T   T   Q   G   Q   N   D   T   P   R   R   S   K   K   R   R   Q   V   D   K   T   A   N   I   E   E   R   Q   Q   R   L   Q   E   I   E
2401/801                                2431/811                                2461/821                                2491/831
GAG AAT CGG CAG GAG CGG ATG AGC AAG CGG CGC CAG CAG CGC TGT CAC CAG ACG CAG AAT TTC TAC AAG AGC CTT CAG CTC CTG CAG GCG GGC AAG CTC TTG AGG GAG GGT GGT GAG GCG
E   N   R   Q   E   R   M   S   K   R   R   Q   Q   R   C   H   Q   T   Q   N   F   Y   K   S   L   Q   L   L   Q   A   G   K   L   L   R   E   G   G   E   A
2521/841                                2551/851                                2581/861                                2611/871
GGA GTG GCC GAG GAT GGC ACC CCA TTC GAG GAC TAC TCG ATA TTC CTC TAC CGC CAG CAG GCG CCC GTA TTC AAT GAT CGC GTC AAG GAC CTC GAG CGA AAG CTT CTG TTT CCC GAT CGC
G   V   A   E   D   G   T   P   F   E   D   Y   S   I   F   L   Y   R   Q   Q   A   P   V   F   N   D   R   V   K   D   L   E   R   K   L   L   F   P   D   R
2641/881                                2671/891                                2701/901                                2731/911
GAA CGG GGA GAT ATT CCG TCG GCA TTG CCG CGC ACG GCG GAC GAG CAG TTC AGC GAT CGG ATT AAA AAC ATG GAG CAG CGG ATG ACG GGA CGT GGT GGC CTT GGT GGT GAC AAG AAG CCC
E   R   G   D   I   P   S   A   L   P   R   T   A   D   E   Q   F   S   D   R   I   K   N   M   E   Q   R   M   T   G   R   G   G   L   G   G   D   K   K   P
2761/921                                2791/931                                2821/941                                2851/951
AAG GAT CTG ATG CGG GCT ATC GGC AAG ATC GAC TCG AAC GAT TGG AAT GTA CGC GAG ATC GAG AAG AAG ATC GAG CTG TCG AAG AAG ACG GAG ATC CAC GGG CCT AAG GGC CGC GAG AAG
K   D   L   M   R   A   I   G   K   I   D   S   N   D   W   N   V   R   E   I   E   K   K   I   E   L   S   K   K   T   E   I   H   G   P   K   G   R   E   K
2881/961                                2911/971                                2941/981                                2971/991
GTT CCC AAG TGG AGC AAG GAG CAG TTC CAG GCG CGA CAG CAC AAG ATG TCC AAG CCG CAA CGC CAG GAT TCG CGT GAG GCG GAA AAG TTC AAG GAC ATC GAC CAG ACT ATC CGC AAC TTG
V   P   K   W   S   K   E   Q   F   Q   A   R   Q   H   K   M   S   K   P   Q   R   Q   D   S   R   E   A   E   K   F   K   D   I   D   Q   T   I   R   N   L
3001/1001                               3031/1011                               3061/1021                               3091/1031
GAC AAG CAG CTA AAG GAG GGC CAC AAT CTG GAT GTG GGC GAG CGG GGA CGC AAC AAA GTG GCC TCC ATT GCC GGT CAG TTT GGC AAA AAG GAT GAG GCC AAT TCG GAT GAG AAG AAC GCG
D   K   Q   L   K   E   G   H   N   L   D   V   G   E   R   G   R   N   K   V   A   S   I   A   G   Q   F   G   K   K   D   E   A   N   S   D   E   K   N   A
3121/1041                               3151/1051                               3181/1061                               3211/1071
GGC AGC AGC AAT GCC ACC ACC AAC ACC AAC AAC ACA GTC ATA CCC AAA TCT AGT TCC AAG GTG GCG CTG GCC TTT AAA AAG CAG GCT GCC TCC GAA AAG TGC CGC TTC TGT AAG CAA ACC
G   S   S   N   A   T   T   N   T   N   N   T   V   I   P   K   S   S   S   K   V   A   L   A   F   K   K   Q   A   A   S   E   K   C   R   F   C   K   Q   T
3241/1081                               3271/1091                               3301/1101                               3331/1111
GTC TAC CTG ATG GAG AAG ACC ACC GTG GAG GGA TTG GTT CTG CAT CGC AAT TGC CTT AAG TGC CAC CAC TGC CAC ACC AAC TTG CGT CTG GGA GGC TAC GCC TTT GAT CGG GAC GAT CCG
V   Y   L   M   E   K   T   T   V   E   G   L   V   L   H   R   N   C   L   K   C   H   H   C   H   T   N   L   R   L   G   G   Y   A   F   D   R   D   D   P
3361/1121                               3391/1131                               3421/1141                               3451/1151
CAG GGC CGA TTT TAC TGC ACC CAA CAC TTC CGG TTG CCA CCC AAA CCG CTG CCG CAG CGC ACC AAC AAA GCC AGG AAA TCC GCT GCC GCT CAA CCC GCC TCG CCT GCT GTA CCA CCA ACT
Q   G   R   F   Y   C   T   Q   H   F   R   L   P   P   K   P   L   P   Q   R   T   N   K   A   R   K   S   A   A   A   Q   P   A   S   P   A   V   P   P   T
3481/1161                               3511/1171                               3541/1181                               3571/1191
GCG GGA TCC GTA CCC ACT GCA GCT GCC ACA TCG GAG CAT ATG GAC ACC ACT CCA CCC AGG GAC CAG GTG GAC CTA CTG GAG ACC TCG CGA GCA AAT GCC TCT GCC GAT GCC ATG TCC GAT
A   G   S   V   P   T   A   A   A   T   S   E   H   M   D   T   T   P   P   R   D   Q   V   D   L   L   E   T   S   R   A   N   A   S   A   D   A   M   S   D
3601/1201                               3631/1211                               3661/1221                               3691/1231
GAT GAA GCC AAT GTT ATC GAT GAG CAC GAA TGG TCT GGT CGC AAC TTC TTG CCC GAG TCC AAC AAC GAT TCC CAA TCA GAG CTA TCC AGT TCA GAT GAG TCG GAT ACG GAA TCG GAT TCG
D   E   A   N   V   I   D   E   H   E   W   S   G   R   N   F   L   P   E   S   N   N   D   S   Q   S   E   L   S   S   S   D   E   S   D   T   E   S   D   S
3721/1241                               3751/1251                               3781/1261                               3811/1271
GAG ATG TTT GAG GAG GCG GAT GAT TCG CCG TTT GGT GCT CAG ACC CTC CAG CTG GCG TCG GAT TGG ATT GGA AAG CAA TAC TGT GAG GAC AGT GAT GAT TCT GAC GAT TTC TAC GAC TCA
E   M   F   E   E   A   D   D   S   P   F   G   A   Q   T   L   Q   L   A   S   D   W   I   G   K   Q   Y   C   E   D   S   D   D   S   D   D   F   Y   D   S
3841/1281                               3871/1291                               3901/1301                               3931/1311
AGT GAA GGT ATT GCG GAT GAC GGC AAA GAT GAC ACC GAG GGT GAG GAA TTC AAG AAG GCC CGC GAA TTG AGG CGC CAG GAA GTT CGC CTG CAG CCG TTG CCC GCC AAT CTG CCC ACA GAT
S   E   G   I   A   D   D   G   K   D   D   T   E   G   E   E   F   K   K   A   R   E   L   R   R   Q   E   V   R   L   Q   P   L   P   A   N   L   P   T   D
3961/1321                               3991/1331                               4021/1341                               4051/1351
ACG GAG ACC GAG GTT CAA ACC GAG TCC GAG AGC ACT TCA CCC GAC GAA GTG GAG CTC AAT TCT GCC ACT GAG ATA TCC ACC GAC TCT GAG TTT GAC AAC GAT GAG ATT ATA CGC CAG GCG
T   E   T   E   V   Q   T   E   S   E   S   T   S   P   D   E   V   E   L   N   S   A   T   E   I   S   T   D   S   E   F   D   N   D   E   I   I   R   Q   A
4081/1361                               4111/1371                               4141/1381                               4171/1391
CCC AAA ATC TTC ATC GAT GAC ACC CAT CTA AGG AAG CCC ACC AAG GTT CAG ATC AAG TCC ACC ATG ATC GGA CCC AAT GCA GCT TCC GCC GGA CTC CAT CAG AAG CAG TTG GCG GCG CGT
P   K   I   F   I   D   D   T   H   L   R   K   P   T   K   V   Q   I   K   S   T   M   I   G   P   N   A   A   S   A   G   L   H   Q   K   Q   L   A   A   R
4201/1401                               4231/1411                               4261/1421                               4291/1431
GAG AAG GGC GGC AGC TAC CTC CAG AAG TAC CAA CCA CAA CCG CCA CTG TCA CAG TTT AAA CCG TTG GTC CAG GTG GAT CCC ACC CTG CTC ATT GGC AGC CAG CGC GCT CCT CTT CAG AAT
E   K   G   G   S   Y   L   Q   K   Y   Q   P   Q   P   P   L   S   Q   F   K   P   L   V   Q   V   D   P   T   L   L   I   G   S   Q   R   A   P   L   Q   N
4321/1441                               4351/1451                               4381/1461                               4411/1471
CCA CGG CCA GGA GAC TAC TTG CTA AAC AAG ACG GCC AGT ACG GAG GGT ATC GCC TCA AAA AAG AGC CTG GAG CTA AAA AAG CGC TAT CTG CTG GGT GAG CCG GCC AAT GGC GAT AAG ATC
P   R   P   G   D   Y   L   L   N   K   T   A   S   T   E   G   I   A   S   K   K   S   L   E   L   K   K   R   Y   L   L   G   E   P   A   N   G   D   K   I
```

FIGURE 15-2

DMICALmedium.txt

```
4441/1481                               4471/1491                               4501/1501                               4531/1511
CAG AAG TCC GGA TCC ACT TCA GTG CTG GAT TCA CGC ATT CGC AGC TTC CAG TCG AAC ATA TCG GAG TGC CAG AAG CTT TTG AAT CCC AGC AGC GAC ATA AGT GCC GGC ATG CGA ACC TTC
Q   K   S   G   S   T   S   V   L   D   S   R   I   R   S   F   Q   S   N   I   S   E   C   Q   K   L   L   N   P   S   S   D   I   S   A   G   M   R   T   F
4561/1521                               4591/1531                               4621/1541                               4651/1551
CTC GAT CGC ACA AAG TTG GGC GAA GGC AGC CAG ACG ACA CCC GGA CAG ACG AAC GAA CTA ATC CGT TCC GCC ACC AGC AAT GTG ATT AAC GAT CTG CGC GTG GAG CTT CGG ATA CAG AAA
L   D   R   T   K   L   G   E   G   S   Q   T   T   P   G   Q   T   N   E   L   I   R   S   A   T   S   N   V   I   N   D   L   R   V   E   L   R   I   Q   K
4681/1561                               4711/1571                               4741/1581                               4771/1591
ACT GAC TCC AGC CAC TCC ACG GAC AAC GAG AAG GAA AAC GTT TTC GTG AAC TGT AAG AAC GAG CTG AAC AAG GGG ATG GAA TAC ACG GAT GCG GTC AAT GCC ACG CTG CTG GAC CAG CTG
T   D   S   S   H   S   T   D   N   E   K   E   N   V   F   V   N   C   K   N   E   L   N   K   G   M   E   Y   T   D   A   V   N   A   T   L   L   D   Q   L
4801/1601                               4831/1611                               4861/1621                               4891/1631
GCC AGA AAA AGT TCA CCC ACC ACG CCG ACG AAT AAG ACG GTG GTC GAG GTT ATT GAC CTG GTT ACA CCT GAG AAG CCA ATT GAC ATT ATC GAT CTA ACG GCA CTG GAA ACG CCG AAA AAG
A   R   K   S   S   P   T   T   P   T   N   K   T   V   V   E   V   I   D   L   V   T   P   E   K   P   I   D   I   I   D   L   T   A   L   E   T   P   K   K
4921/1641                               4951/1651                               4981/1661                               5011/1671
CAG TTG GTC GAT GGT AGC GCC ATG GAT GTA GAT GAA CGC CTC ACA CCC GAT AGC AAC AAA ATC AGC GAA CTG CAG CAG GAA GTG AAG GAG GAA CCC AAG CCG GAT GTC TCT AGG GAT GTG
Q   L   V   D   G   S   A   M   D   V   D   E   R   L   T   P   D   S   N   K   I   S   E   L   Q   Q   E   V   K   E   E   P   K   P   D   V   S   R   D   V
5041/1681                               5071/1691                               5101/1701                               5131/1711
AAA GAA TGC ATA CCA GAT ATA CTG GGA CAC ATT AAG GAG GGA ACG GGA TCG AAG GAG CCA GGT GGA GAG GAC CAA CAG AGC CTG CTG GAG CAG TCG GAC GAA GAG AAG CGC GAC TCA CCG
K   E   C   I   P   D   I   L   G   H   I   K   E   G   T   G   S   K   E   P   G   G   E   D   Q   Q   S   L   L   E   Q   S   D   E   E   K   R   D   S   P
5161/1721                               5191/1731                               5221/1741                               5251/1751
GAA AAG GAT GTG GCC GAA CAT GAG CTT TAT GAA CCG GAC AGT GTG CAG ATC CAG GTG CCC AAT ATC CCA TGG GAA AAA AGC AAG CCG GAG GTC ATG TCT ACC ACC GGC AGC AGT GGC TCC
E   K   D   V   A   E   H   E   L   Y   E   P   D   S   V   Q   I   Q   V   P   N   I   P   W   E   K   S   K   P   E   V   M   S   T   T   G   S   S   G   S
5281/1761                               5311/1771                               5341/1781                               5371/1791
ATC TGC TCA AGC TCA GAC TCT TCT AGT ATT GAA GAC ATC CAG CAC TAC ATT TTG GAG TCC ACA ACT AGT CCA GAT ACT CAG ACA GTT GGC GGA AAG CAC AAT GTG CCC CGT TTG GAG GTG
I   C   S   S   S   D   S   S   S   I   E   D   I   Q   H   Y   I   L   E   S   T   T - S   P   D   T   Q   T   V   G   G   K   H   N   V   P   R   L   E   V
5401/1801                               5431/1811                               5461/1821                               5491/1831
CAC GAC ACA AGT GGT GCC CTG ATG CAG GTG GAC AGC CTG ATG ATT GTG AAC GGA AAG TAT ATT GGG GAT CCC GAG GAT GTC AAG TTC TTG GAT ATG CCG GCC AAT GTT ATT GTT CCG CCA
H   D   T   S   G   A   L   M   Q   V   D   S   L   M   I   V   N   G   K   Y   I   G   D   P   E   D   V   K   F   L   D   M   P   A   N   V   I   V   P   P
5521/1841                               5551/1851                               5581/1861                               5611/1871
GCA CCG GCG CTT AAA ACG AAT GAG CTG GAT ATG GAG GAT GAC CAA GAG GCG GAG GCG GAA CCA GTA ACT GCT ACT CCG GAG CCG GTG GAA TGT ACG GTC ATC GAG GCT GAG CGC CGT GTT
A   P   A   L   K   T   N   E   L   D   M   E   D   D   Q   E   A   E   A   E   P   V   T   A   T   P   E   P   V   E   C   T   V   I   E   A   E   R   R   V
5641/1881                               5671/1891                               5701/1901                               5731/1911
ACT GCT CCC CCT CCT TTG CCA GAG ATG GGT CCA CCC AAA CTG AAG TTC GAT AGC AAA AAT GAG AAC AAG ATC GAG AGC TTG AAG AAT CTT CCG TTG ATC GTA GAG AGC AAT GTG GAG CAC
T   A   P   P   P   L   P   E   M   G   P   P   K   L   K   F   D   S   K   N   E   N   K   I   E   S   L   K   N   L   P   L   I   V   E   S   N   V   E   H
5761/1921                               5791/1931                               5821/1941                               5851/1951
AGT CAG GCA GTG AAA CCC ATT ACT CTT AAC TTA AGC AAT CTG GCC AGG ACG CCG GAT ACA CCA ACC ACG CCC ACG GCG CAC GAT AGC GAT AAA ACA CCC ACT GGG GAA ATT CTC TCG CGA
S   Q   A   V   K   P   I   T   L   N   L   S   N   L   A   R   T   P   D   T   P   T   T   P   T   A   H   D   S   D   K   T   P   T   G   E   I   L   S   R
5881/1961                               5911/1971                               5941/1981                               5971/1991
GGA TCT GAC TCA GAA ACC GAG CAC ACT GGC ACT GGT CAG GTA CTA ACG GAG ACG GAA CTC TCC GAC TGG ACG GCC GAC GAC TGT ATC TCG GAG AAC TTT GTT GAC TTG GAG TTC GCG CTT
G   S   D   S   E   T   E   H   T   G   T   G   Q   V   L   T   E   T   E   L   S   D   W   T   A   D   D   C   I   S   E   N   F   V   D   L   E   F   A   L
6001/2001                               6031/2011                               6061/2021                               6091/2031
AAC TCT AAC AAG GGT ACG ATA AAA CGG CGC AAG GAT CGA CGA CGC AGT GGA GCA AGC AAA CTT CCC AGT GGC AAC GAG GTA ATC CAC GAG CTG GCC AGG CAG GCG CCA GTG GTG CAA ATG
N   S   N   K   G   T   I   K   R   R   K   D   R   R   R   S   G   A   S   K   L   P   S   G   N   E   V   I   H   E   L   A   R   Q   A   P   V   V   Q   M
6121/2041                               6151/2051                               6181/2061                               6211/2071
GAT GGA ATT CTT AGT GCC ATC GAC ATT GAT GAC ATA GAG TTC ATG GAC ACG GGT TCG GAG GGT TCT TGT GCT GAA GCT TAT CCC GCA ACA AAT ACA GCT CTC ATT CAG AAT AGA GGT TAC
D   G   I   L   S   A   I   D   I   D   D   I   E   F   M   D   T   G   S   E   G   S   C   A   E   A   Y   P   A   T   N   T   A   L   I   Q   N   R   G   Y
6241/2081                               6271/2091                               6301/2101                               6331/2111
ATG GAG TAC ATC GAG GCG GAG CCG AAA AAG ACG ACC CGC AAG GCA GCT CCA CCA TCC AGT TAC CCA GGA AAT TTA CCG CCT TTA ATG ACG AAG CGG GAC GAG AAA CTG GGC GTT GAT TAC
M   E   Y   I   E   A   E   P   K   K   T   T   R   K   A   A   P   P   S   S   Y   P   G   N   L   P   P   L   M   T   K   R   D   E   K   L   G   V   D   Y
6361/2121                               6391/2131                               6421/2141                               6451/2151
ATT GAG CAG GGG GCG TAC ATA ATG CAC GAT GAT GCA AAG ACG CCT GTG AAT GAG GTG GCT CCT GCC ATG ACC CAG TCG CTA ACT GAC TCA ATC ACG CTC AAT GAA CTG GAT GAT GAC AGT
I   E   Q   G   A   Y   I   M   H   D   D   A   K   T   P   V   N   E   V   A   P   A   M   T   Q   S   L   T   D   S   I   T   L   N   E   L   D   D   D   S
6481/2161                               6511/2171                               6541/2181                               6571/2191
ATG ATA ATA TCC CAA ACC CAG CCA ACG ACA ACG GAG GAA AGT GAG GCA CTG ACG GTG GTC ACC AGT CCA CTT GAC ACG TCC TCG CCC AGG GTT CTC GAT CAA TTT GCA TCC ATG TTG GCG
M   I   I   S   Q   T   Q   P   T   T   T   E   E   S   E   A   L   T   V   V   T   S   P   L   D   T   S   S   P   R   V   L   D   Q   F   A   S   M   L   A
6601/2201                               6631/2211                               6661/2221                               6691/2231
GCG GGA AAA GGT GAC TCC ACA CCC AGT AGC TCA GAG CAA CAA CCA AAG ACG TCT ACG GTG ACG AGC AGC AGC ACT GGG CCC AAC TCC TCG ACA ACA GGA AAC GTC TCG AAG GAG CCG CAG
```

FIGURE 15-3

DMICALmedium.txt

```
A   G   K   G   D   S   T   P   S   S   S   E   Q   Q   P   K   T   S   T   V   T   S   S   S   T   G   P   N   S   S   T   T   G   N   V   S   K   E   P   Q
6721/2241                               6751/2251                               6781/2261                               6811/2271
GAG GAG GAC CTG CAA ATC CAG TTT GAG TAT GTT CGA GCA CTG CAG CAG CGG ATA TCG CAG ATC AGC ACC CAA CGG CGT AAG AGC TCT AAG GGA GAG GCA CCT AAC CTG CAG CTA AAC AGT
E   E   D   L   Q   I   Q   F   E   Y   V   R   A   L   Q   Q   R   I   S   Q   I   S   T   Q   R   R   K   S   S   K   G   E   A   P   N   L   Q   L   N   S
6841/2281                               6871/2291                               6901/2301                               6931/2311
AGC GCA CCT GTG ATA GAA TCA GCC GAG GAT CCG GCC AAG CCC GCA GAG GAG CCT CTG GTC TCA ATG CGA CCG CGG ACC ACC AGC ATT TCC GGA AAG GTA CCG GAG ATA CCC ACA CTT AGC
S   A   P   V   I   E   S   A   E   D   P   A   K   P   A   E   E   P   L   V   S   M   R   P   R   T   T   S   I   S   G   K   V   P   E   I   P   T   L   S
6961/2321                               6991/2331                               7021/2341                               7051/2351
AGC AAG CTG GAA GAG ATA ACC AAA GAA CGC ACT AAG CAA AAG GAT CTG ATT CAC GAC CTA GTC ATG GAC AAG TTG CAG TCG AAG AAG CAG CTA AAC GCT GAG AAG CGT CTG CAC CGG AGT
S   K   L   E   E   I   T   K   E   R   T   K   Q   K   D   L   I   H   D   L   V   M   D   K   L   Q   S   K   K   Q   L   N   A   E   K   R   L   H   R   S
7081/2361                               7111/2371                               7141/2381                               7171/2391
CGA CAG CGC AGT TTG CTG ACC AGT GGC TAT GCC AGT GGC TCC AGC CTT AGT CCG ACG CCC AAG CTG GCT GCT GCT TGC AGT CCG CAG GAT TCC AAC TGC TCT AGC CAA GCG CAC TAC CAC
R   Q   R   S   L   L   T   S   G   Y   A   S   G   S   S   L   S   P   T   P   K   L   A   A   A   C   S   P   Q   D   S   N   C   S   S   Q   A   H   Y   H
7201/2401                               7231/2411                               7261/2421                               7291/2431
GCC TCC ACG GCG GAG GAG GCC CCG AAG CCG CCG GCG GAA AGG CCG TTG CAG AAG TCC GCC ACG TCC ACC TAT GTG TCG CCT TAT CGC ACT GTC CAA GCG CCC ACA CGT AGT GCT GAT CTC
A   S   T   A   E   E   A   P   K   P   P   A   E   R   P   L   Q   K   S   A   T   S   T   Y   V   S   P   Y   R   T   V   Q   A   P   T   R   S   A   D   L
7321/2441                               7351/2451                               7381/2461                               7411/2471
TAT AAG CCG CGC CCC TTC AGC GAA CAC ATC GAT TCG AAC GCT CTG GCG GGT TAC AAG CTC GGC AAG ACG GCC TCG TTT AAT GGC GGC AAG TTG GGC GAC TTT GCG AAA CCC ATT GCC CCG
Y   K   P   R   P   F   S   E   H   I   D   S   N   A   L   A   G   Y   K   L   G   K   T   A   S   F   N   G   G   K   L   G   D   F   A   K   P   I   A   P
7441/2481                               7471/2491                               7501/2501                               7531/2511
GCG AGA GTT AAC CGA GGA GGA GGT GTC GCG ACC GCG GAT ATA GCC AAT ATT TCC GCG TCG ACG GAG AAC CTA AGA AGC GAG GCC AGG GCC AGG GCT CGT CTT AAG TCT AAC ACA GAG CTG
A   R   V   N   R   G   G   G   V   A   T   A   D   I   A   N   I   S   A   S   T   E   N   L   R   S   E   A   R   A   R   A   R   L   K   S   N   T   E   L
7561/2521                               7591/2531                               7621/2541                               7651/2551
GGC CTT AGT CCC GAG GAA AAG ATG CAG CTA ATA CGT TCA AGA TTG CAC TAC GAC CAA AAC AGA TCT CTG AAG CCG AAG CAA CTG GAG GAG ATG CCA TCC GGG GAT CTG GCG GCA CGT GCC
G   L   S   P   E   E   K   M   Q   L   I   R   S   R   L   H   Y   D   Q   N   R   S   L   K   P   K   Q   L   E   E   M   P   S   G   D   L   A   A   R   A
7681/2561                               7711/2571                               7741/2581                               7771/2591
CGC AAA ATG AGT GCC TCG AAG AGC GTC AAT GAT CTG GCC TAC ATG GTG GGA CAG CAG CAG CAG CAG CAG GTT GAG AAG GAT GCC GTG CTC CAA GCC AAG GCG GCT GAC TTT ACA TCC GAT
R   K   M   S   A   S   K   S   V   N   D   L   A   Y   M   V   G   Q   Q   Q   Q   Q   Q   V   E   K   D   A   V   L   Q   A   K   A   A   D   F   T   S   D
7801/2601                               7831/2611                               7861/2621                               7891/2631
CCC AAT TTG GCG TCC GGT GGT CAG GAG AAG GCA GGC AAA ACT AAG TCC GGA CGC AGG CCA AAG GAT CCG GAG CGG CGT AAG AGT CTC ATA CAG TCG TTG TCC AGC TTC TTC CAA AAG GGA
P   N   L   A   S   G   G   Q   E   K   A   G   K   T   K   S   G   R   R   P   K   D   P   E   R   R   K   S   L   I   Q   S   L   S   S   F   F   Q   K   G
7921/2641                               7951/2651                               7981/2661                               8011/2671
TCT GGA TCC GCG GCC TCC AGT TCC AAG GAG CAG GGC GGC GCT GTG GCT GCC GTC CAC TCT GAA CAG TCA GAG CGA CCA GGC ACC AGC AGC AGC GGC ACG CCC ACA ATA TCG GAT GCG GCG
S   G   S   A   A   S   S   S   K   E   Q   G   G   A   V   A   A   V   H   S   E   Q   S   E   R   P   G   T   S   S   S   G   T   P   T   I   S   D   A   A
8041/2681                               8071/2691                               8101/2701                               8131/2711
GGT GGA GGC GGA GGA GGA GGT GGC GTC TTC AGC AGA TTC CGC ATC TCG CCC AAG TCC AAG GAG AAG TCA AAG TCT TGC TTT GAT CTG AGG AAT TTC GGT TTT GGT GAC AAG GAT ATG CTG
G   G   G   G   G   G   G   G   V   F   S   R   F   R   I   S   P   K   S   K   E   K   S   K   S   C   F   D   L   R   N   F   G   F   G   D   K   D   M   L
8161/2721                               8191/2731                               8221/2741                               8251/2751
GTC TGC AAT GCA GCA TCT CCA GCA GGA GCC ACA TCC GCA TCA CAG AAA AAT CAC TCG CAA GAG TAT CTG AAC ACC ACG AAC AAC AGT CGC TAT CGA AAG CAA ACG AAC ACT GCG AAA CCG
V   C   N   A   A   S   P   A   G   A   T   S   A   S   Q   K   N   H   S   Q   E   Y   L   N   T   T   N   N   S   R   Y   R   K   Q   T   N   T   A   K   P
8281/2761                               8311/2771                               8341/2781                               8371/2791
AAA CCC GAA TCG TTC TCT TCA TCC AGT CCG CAG CTC TAT ATA CAC AAG CCC CAC CAC CTG GCC GCA GCT CAT CCC AGT GCC CTG GAC GAC CAG ACA CCA CCA CCC ATA CCG CCT CTT CCA
K   P   E   S   F   S   S   S   S   P   Q   L   Y   I   H   K   P   H   H   L   A   A   A   H   P   S   A   L   D   D   Q   T   P   P   P   I   P   P   L   P
8401/2801                               8431/2811                               8461/2821                               8491/2831
CTG AAT TAT CAG AGA TCC GAT GAT GAG AGC TAC GCT AAC GAG ACA CGA GAG CAT AAG AAG CAA CGT GCC ATA TCG AAG GCT TCA CGA CAA GCT GAG CTC AAG CGA TTG CGA ATC GCT CAA
L   N   Y   Q   R   S   D   D   E   S   Y   A   N   E   T   R   E   H   K   K   Q   R   A   I   S   K   A   S   R   Q   A   E   L   K   R   L   R   I   A   Q
8521/2841                               8551/2851                               8581/2861                               8611/2871
GAG ATT CAG CGG GAA CAG GAG GAG ATC GAG GTG CAA CTG AAG GAT CTG GAG GCA CGC GGC GTG CTT ATT GAG AAG GCC TTG CGA GGC GAG GCG CAG AAT ATT GAA AAC CTG GAT GCG ACA
E   I   Q   R   E   Q   E   E   I   E   V   Q   L   K   D   L   E   A   R   G   V   L   I   E   K   A   L   R   G   E   A   Q   N   I   E   N   L   D   A   T
8641/2881                               8671/2891                               8701/2901                               8731/2911
AAG GAC AAC GAC GAG AAG CTA CTT AAG GAA CTT TTG GAG ATT TGG CGC AAC ATC ACA GCA CTC AAG AAA CGC GAT GAG GAA CTG ACT ATA AGG CAA CAG GAA CTG CAA CTG GAG TAT CGG
K   D   N   D   E   K   L   L   K   E   L   L   E   I   W   R   N   I   T   A   L   K   K   R   D   E   E   L   T   I   R   Q   Q   E   L   Q   L   E   Y   R
8761/2921                               8791/2931                               8821/2941                               8851/2951
CAT GCC CAG CTG AAG GAA GAG CTC AAT CTG CGC TTG TCC TGC AAC AAA CTG GAC AAA AGC TCT GCC GAT GTG GCC GCC GAG GGA GCA ATT CTC AAC GAG ATG CTG GAA ATT GTC GCC AAG
H   A   Q   L   K   E   E   L   N   L   R   L   S   C   N   K   L   D   K   S   S   A   D   V   A   A   E   G   A   I   L   N   E   M   L   E   I   V   A   K
8881/2961                               8911/2971                               8941/2981                               8971/2991
```

FIGURE 15-4

DMICALmedium.txt

```
CGA GCC GCC CTA CGA CCC ACA GCC TCC CAG CTC GAC CTC ACG GCA GCG GGA TCA GCA TCC ACG TCC GCC GAG GCA ACG GGC ATT AAG CTG ACG GGA CAA CCG CAT GAC CAC GAA GAA TCG
R   A   A   L   R   P   T   A   S   Q   L   D   L   T   A   A   G   S   A   S   T   S   A   E   A   T   G   I   K   L   T   G   Q   P   H   D   H   E   E   S
9001/3001
ATC ATT TGA
I   I   *
ATAGCCAACTTTCCCTGCAGATTGCTGGTCTGCGAAGTGGTGCATGGACTGTTGAGCCTAAATACTTAAGATGAAACAAGGGAACAGGGGAGAGCGTACCAAAAAGGCAACTGAATATACAGAATTAACAATACTAAAACCCGTGAGCTGTAACCGAAAC
GAGCAACTAACTATTCAAAGTGGATAAGCTGGCTAAGAACAATTTCAATATTTATATAAGTAGAGAGCCTAACTAATATGGGAGAAGCGCGCAAGGTTAATGTAAAAACAGGCAGAGATGAAATCAAAACCAAAAAACTAATTTCGCAATTTAAAAGTAG
GGATTACGAATGTTTTTGCGAATACCTTTAGTTTCGTATTCGACCTAATTGTACGTTTTGTTAGTCGTTGCCACTGAAATCTGTCTGATTTTTGTATGTCCTATGCAATATGTTGTGTTTTTATTTAATGTAATGTAGTGAGCGTGTGCAAAGGGTATGT
GATTTTGAGCAAATGTAAATGGTGTATTTGAGTTTTCTTAGGATATTTTCACTTAGCGATTGTTTGTCGATTTCTATGTAAACCGTTTGAGTGTATTATTATTTATTTGACAAAGCACCTTCCGCACATCCTCCCGATCAGCACACCCATGTATGTATTA
TGTCATTTGCCCCCAATCGAATTCAAAGTGAACACAAAGAGATGCACACAAATATTTTCTAAATGCGAGAGCCTACACATAGCGATGTAAGCGAGAAAGCAACCATTAACTGTATACTAACGAAATCATTTAAATTATTGCGTAACTTTATTTATAACCT
GCATATATAAAAACATATTTATCGAATTTATTATGTAACTTTTCGATTTATTTTGAATTGAATAAAGATTTTATGAAACCTAA
```

FIGURE 15-5

DMICALshort.txt

```
GCTCTGCATCTCGCAACGAATCGTTCCGAATAAAAAATTAAATAAATAAACACAAAAAATTGTGATATAAGCGATAAATATTGTGATAATTCGGACAAACGTGTCAATTGCATTGCATCAGA
1/1                                     31/11                                   61/21                                   91/31
ATG AGC CGC CAA CAC CAG CGG CAC CAC CAG CAG CAT CAC CAC CTG CCG CCG CAC CAG CAA CCG CAG CAG CAG ATG CCG CAA CAA CAG CAG CAG CTG ACG GCG CAG CAG CAG CAA CAA CAG
M   S   R   Q   H   Q   R   H   H   Q   Q   H   H   H   L   P   P   H   Q   Q   P   Q   Q   Q   M   P   Q   Q   Q   Q   Q   L   T   A   Q   Q   Q   Q   Q   Q
121/41                                  151/51                                  181/61                                  211/71
CAG CTG CTG ATG GCG GAG CAC GCG GCG GCC GCG GAG GCG GCG GAG CTA TTC GAC CTG CTG TGC GTG GCC ACA ACG ATG CGC CAG ATC CTG GCG CTC CAT CGG GCC ATG TGC GAG GCT GTG
Q   L   L   M   A   E   H   A   A   A   A   E   A   A   E   L   F   D   L   L   C   V   A   T   T   M   R   Q   I   L   A   L   H   R   A   M   C   E   A   V
241/81                                  271/91                                  301/101                                 331/111
GGA TTG AGA CCC TCG CCT CTG AAC GAC TTC TAC CCA CGG CTA AAG GCC AAG GTG CGT TCG TGG AAG GCG CAG GCC CTG TGG AAG AAG TTC GAC GCC AGA GCT GCC CAT AGA GTC TAC GGC
G   L   R   P   S   P   L   N   D   F   Y   P   R   L   K   A   K   V   R   S   W   K   A   Q   A   L   W   K   K   F   D   A   R   A   A   H   R   V   Y   G
361/121                                 391/131                                 421/141                                 451/151
AAG GGA GCT GCC TGT ACT GGC ACA CGC GTC CTG GTC ATC GGA GCA GGG CCC TGT GGA CTG CGC ACC GCC ATC GAG GCC CAA CTG CTG GGC GCC AAG GTG GTG GTG CTG GAG AAA CGC GAT
K   G   A   A   C   T   G   T   R   V   L   V   I   G   A   G   P   C   G   L   R   T   A   I   E   A   Q   L   L   G   A   K   V   V   V   L   E   K   R   D
481/161                                 511/171                                 541/181                                 571/191
CGC ATC ACC CGG AAC AAT GTG CTC CAT CTG TGG CCA TTC GTC ATC ACG GAT CTG CGC AAC TTG GGC GCA AAG AAG TTC TAC GGC AAG TTT TGC GCC GGC TCC ATC GAT CAC ATC TCC ATT
R   I   T   R   N   N   V   L   H   L   W   P   F   V   I   T   D   L   R   N   L   G   A   K   K   F   Y   G   K   F   C   A   G   S   I   D   H   I   S   I
601/201                                 631/211                                 661/221                                 691/231
CGG CAG CTG CAG TGC ATG CTG CTC AAG GTG GCG CTG CTC CTG GGC GTA GAG ATC CAC GAG GGA GTC AGT TTT GAT CAC GCT GTA GAG CCC TCT GGC GAT GGC GGC GGA TGG AGG GCA GCT
R   Q   L   Q   C   M   L   L   K   V   A   L   L   L   G   V   E   I   H   E   G   V   S   F   D   H   A   V   E   P   S   G   D   G   G   G   W   R   A   A
721/241                                 751/251                                 781/261                                 811/271
GTT ACT CCC GCA GAT CAT CCT GTA TCT CAC TAC GAA TTC GAT GTG TTG ATC GGA GCG GAT GGC AAG CGG AAT ATG CTG GAC TTT AGG AGG AAG GAG TTC CGC GGG AAG CTG GCC ATC GCT
V   T   P   A   D   H   P   V   S   H   Y   E   F   D   V   L   I   G   A   D   G   K   R   N   M   L   D   F   R   R   K   E   F   R   G   K   L   A   I   A
841/281                                 871/291                                 901/301                                 931/311
ATT ACA GCG AAC TTT ATC AAC AAG AAG ACG GAG GCG GAG GCT AAA GTA GAG GAG ATC AGT GGG GTG GCT TTC ATC TTC AAC CAG GCC TTC TTC AAG GAG CTG TAC GGG AAG ACG GGC ATC
I   T   A   N   F   I   N   K   K   T   E   A   E   A   K   V   E   E   I   S   G   V   A   F   I   F   N   Q   A   F   F   K   E   L   Y   G   K   T   G   I
961/321                                 991/331                                 1021/341                                1051/351
GAC CTG GAA AAC ATC GTC TAC TAC AAG GAC GAG ACG CAC TAC TTC GTG ATG ACG GCC AAG AAG CAC AGT CTA ATT GAC AAG GGC GTT ATT ATC GAG GAT ATG GCC GAT CCC GGC GAG CTT
D   L   E   N   I   V   Y   Y   K   D   E   T   H   Y   F   V   M   T   A   K   K   H   S   L   I   D   K   G   V   I   I   E   D   M   A   D   P   G   E   L
1081/361                                1111/371                                1141/381                                1171/391
CTC GCC CCA GCC AAT GTG GAT ACA CAA AAG CTG CAC GAC TAT GCA CGC GAG GCT GCG GAG TTC TCC ACC CAA TAC CAA ATG CCA AAC CTG GAG TTC GCT GTT AAT CAC TAC GGC AAA CCA
L   A   P   A   N   V   D   T   Q   K   L   H   D   Y   A   R   E   A   A   E   F   S   T   Q   Y   Q   M   P   N   L   E   F   A   V   N   H   Y   G   K   P
1201/401                                1231/411                                1261/421                                1291/431
GAT GTG GCC ATG TTC GAC TTC ACA TCG ATG TTT GCC GCC GAG ATG TCC TGT CGG GTG ATT GTG CGC AAA GGA GCT CGC CTG ATG CAG TGC CTC GTG GGT GAC AGT CTG CTC GAG CCG TTT
D   V   A   M   F   D   F   T   S   M   F   A   A   E   M   S   C   R   V   I   V   R   K   G   A   R   L   M   Q   C   L   V   G   D   S   L   L   E   P   F
1321/441                                1351/451                                1381/461                                1411/471
TGG CCC ACT GGA TCG GGT TGT GCC CGT GGA TTC TTA TCC AGC ATG GAT GCT GCC TAT GCC ATC AAG CTT TGG TCC AAC CCG CAG AAC AGC ACA CTT GGC GTT CTG GCG CAG CGC GAA AGC
W   P   T   G   S   G   C   A   R   G   F   L   S   S   M   D   A   A   Y   A   I   K   L   W   S   N   P   Q   N   S   T   L   G   V   L   A   Q   R   E   S
1441/481                                1471/491                                1501/501                                1531/511
ATC TAC CGG CTG CTT AAC CAG ACC ACG CCG GAC ACC CTG CAG CGG GAC ATC AGT GCC TAT ACC GTG GAT CCG GCC ACG CGC TAT CCG AAT CTG AAC AGG GAG TCG GTC AAT AGC TGG CAG
I   Y   R   L   L   N   Q   T   T   P   D   T   L   Q   R   D   I   S   A   Y   T   V   D   P   A   T   R   Y   P   N   L   N   R   E   S   V   N   S   W   Q
1561/521                                1591/531                                1621/541                                1651/551
GTC AAA CAT CTG GTC GAC ACG GAC GAC CCG TCC ATT CTG GAG CAG ACC TTC ATG GAC ACG CAT GCT CTG CAG ACC CCG CAT TTG GAC ACA CCG GGC AGA CGC AAG CGA CGC AGT GGA GAC
V   K   H   L   V   D   T   D   D   P   S   I   L   E   Q   T   F   M   D   T   H   A   L   Q   T   P   H   L   D   T   P   G   R   R   K   R   R   S   G   D
1681/561                                1711/571                                1741/581                                1771/591
TTG CTG CCC CAG GGT GCC ACG TTG CTG AGA TGG ATA AGT GCC CAG CTG CAT TCC TAT CAG TTT ATT CCC GAA CTC AAG GAG GCT TCG GAT GTG TTC CGG AAT GGA CGC GTT CTG TGT GCG
L   L   P   Q   G   A   T   L   L   R   W   I   S   A   Q   L   H   S   Y   Q   F   I   P   E   L   K   E   A   S   D   V   F   R   N   G   R   V   L   C   A
1801/601                                1831/611                                1861/621                                1891/631
CTT ATC AAT CGC TAT CGT CCT GAT CTC ATC GAC TAC GCT GCC ACC AAG GAC ATG AGT CCC GTG GAG TGC AAT GAG CTG TCA TTC GCC GTC CTA GAG CGC GAA CTC CAC ATC GAT CGC GTC
L   I   N   R   Y   R   P   D   L   I   D   Y   A   A   T   K   D   M   S   P   V   E   C   N   E   L   S   F   A   V   L   E   R   E   L   H   I   D   R   V
1921/641                                1951/651                                1981/661                                2011/671
ATG AGT GCC AAA CAG TCG CTG GAC TTG ACC GAG CTG GAG TCG CGA ATC TGG CTC AAC TAT TTG GAC CAG ATC TGC GAC TTG TTT CGC GGC GAG ATC CCC CAT ATC AAG CAC CCC AAG ATG
M   S   A   K   Q   S   L   D   L   T   E   L   E   S   R   I   W   L   N   Y   L   D   Q   I   C   D   L   F   R   G   E   I   P   H   I   K   H   P   K   M
2041/681                                2071/691                                2101/701                                2131/711
GAC TTT AGC GAT TTG CGC CAG AAG TAT CGT ATC AAC CAT ACG CAT GCC CAA CCC GAC TTC TCC AAG CTG CTG GCA ACG AAA CCC AAG GCC AAG TCG CCG ATG CAG GAT GCT GTG GAC ATA
D   F   S   D   L   R   Q   K   Y   R   I   N   H   T   H   A   Q   P   D   F   S   K   L   L   A   T   K   P   K   A   K   S   P   M   Q   D   A   V   D   I
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 16-1

```
DMICALshort.txt
CCC ACG ACA GTG CAG CGG CGC TCG GTG CTC GAG GAG GAG CGA GCC AAG CGG CAG CGT CGC CAC GAG CAG CTT CTT AAC ATC GGT GGA GGG GCA GCA GGA GCC GCC GCC GGA GTT GCC GGA
P   T   T   V   Q   R   R   S   V   L   E   E   E   R   A   K   R   Q   R   R   H   E   Q   L   L   N   I   G   G   G   A   A   G   A   A   A   G   V   A   G
2281/761                                2311/771                                2341/781                                2371/791
AGC GGG ACA GGA ACC ACA ACG CAG GGT CAA AAC GAT ACG CCA CGC CGG TCC AAG AAG CGC CGT CAG GTT GAC AAA ACC GCC AAT ATT AGT TCC AAG GTG GCG CTG GCC TTT AAA AAG CAG
S   G   T   G   T   T   T   Q   G   Q   N   D   T   P   R   R   S   K   K   R   R   Q   V   D   K   T   A   N   I   S   S   K   V   A   L   A   F   K   K   Q
2401/801                                2431/811                                2461/821                                2491/831
GCT GCC TCC GAA AAG TGC CGC TTC TGT AAG CAA ACC GTC TAC CTG ATG GAG AAG ACC ACC GTG GAG GGA TTG GTT CTG CAT CGC AAT TGC CTT AAG TGC CAC CAC TGC CAC ACC AAC TTG
A   A   S   E   K   C   R   F   C   K   Q   T   V   Y   L   M   E   K   T   T   V   E   G   L   V   L   H   R   N   C   L   K   C   H   H   C   H   T   N   L
2521/841                                2551/851                                2581/861                                2611/871
CGT CTG GGA GGC TAC GCC TTT GAT CGG GAC GAT CCG CAG GGC CGA TTT TAC TGC ACC CAA CAC TTC CGG TTG CCA CCC AAA CCG CTG CCG CAG CGC ACC AAC AAA GCC AGG AAA TCC GCT
R   L   G   G   Y   A   F   D   R   D   D   P   Q   G   R   F   Y   C   T   Q   H   F   R   L   P   P   K   P   L   P   Q   R   T   N   K   A   R   K   S   A
2641/881                                2671/891                                2701/901                                2731/911
GCC GCT CAA CCC GCC TCG CCT GCT GTA CCA CCA ACT GCG GGA TCC GTA CCC ACT GCA GCT GCC ACA TCG GAG CAT ATG GAC ACC ACT CCA CCC AGG GAC CAG GTG GAC CTA CTG GAG ACC
A   A   Q   P   A   S   P   A   V   P   P   T   A   G   S   V   P   T   A   A   A   T   S   E   H   M   D   T   T   P   P   R   D   Q   V   D   L   L   E   T
2761/921                                2791/931                                2821/941                                2851/951
TCG CGA GCA AAT GCC TCT GCC GAT GCC ATG TCC GAT GAT GAA GCC AAT GTT ATC GAT GAG CAC GAA TGG TCT GGT CGC AAC TTC TTG CCC GAG TCC AAC AAC GAT TCC CAA TCA GAG CTA
S   R   A   N   A   S   A   D   A   M   S   D   D   E   A   N   V   I   D   E   H   E   W   S   G   R   N   F   L   P   E   S   N   N   D   S   Q   S   E   L
2881/961                                2911/971                                2941/981                                2971/991
TCC AGT TCA GAT GAG TCG GAT ACG GAA TCG GAT TCG GAG ATG TTT GAG GAG GCG GAT GAT TCG CCG TTT GGT GCT CAG ACC CTC CAG CTG GCG TCG GAT TGG ATT GGA AAG CAA TAC TGT
S   S   S   D   E   S   D   T   E   S   D   S   E   M   F   E   E   A   D   D   S   P   F   G   A   Q   T   L   Q   L   A   S   D   W   I   G   K   Q   Y   C
3001/1001                               3031/1011                               3061/1021                               3091/1031
GAG GAC AGT GAT GAT TCT GAC GAT TTC TAC GAC TCA AGT GAA GGT ATT GCG GAT GAC GGC AAA GAT GAC ACC GAG GGT GAG GAA TTC AAG AAG GCC CGC GAA TTG AGG CGC CAG GAA GTT
E   D   S   D   D   S   D   D   F   Y   D   S   S   E   G   I   A   D   D   G   K   D   D   T   E   G   E   E   F   K   K   A   R   E   L   R   R   Q   E   V
3121/1041                               3151/1051                               3181/1061                               3211/1071
CGC CTG CAG CCG TTG CCC GCC AAT CTG CCC ACA GAT ACG GAG ACC GAG GTT CAA ACC GAG TCC GAG AGC ACT TCA CCC GAC GAA GTG GAG CTC AAT TCT GCC ACT GAG ATA TCC ACC GAC
R   L   Q   P   L   P   A   N   L   P   T   D   T   E   T   E   V   Q   T   E   S   E   S   T   S   P   D   E   V   E   L   N   S   A   T   E   I   S   T   D
3241/1081                               3271/1091                               3301/1101                               3331/1111
TCT GAG TTT GAC AAC GAT GAG ATT ATA CGC CAG GCG CCC AAA ATC TTC ATC GAT GAC ACC CAT CTA AGG AAG CCC ACC AAG GTT CAG ATC AAG TCC ACC ATG ATC GGA CCC AAT GCA GCT
S   E   F   D   N   D   E   I   I   R   Q   A   P   K   I   F   I   D   D   T   H   L   R   K   P   T   K   V   Q   I   K   S   T   M   I   G   P   N   A   A
3361/1121                               3391/1131                               3421/1141                               3451/1151
TCC GCC GGA CTC CAT CAG AAG CAG TTG GCG GCG CGT GAG AAG GGC GGC AGC TAC CTC CAG AAG TAC CAA CCA CAA CCG CCA CTG TCA CAG TTT AAA CCG TTG GTC CAG GTG GAT CCC ACC
S   A   G   L   H   Q   K   Q   L   A   A   R   E   K   G   G   S   Y   L   Q   K   Y   Q   P   Q   P   P   L   S   Q   F   K   P   L   V   Q   V   D   P   T
3481/1161                               3511/1171                               3541/1181                               3571/1191
CTG CTC ATT GGC AGC CAG CGC GCT CCT CTT CAG AAT CCA CGG CCA GGA GAC TAC TTG CTA AAC AAG ACG GCC AGT ACG GAG GGT ATC GCC TCA AAA AAG AGC CTG GAG CTA AAA AAG CGC
L   L   I   G   S   Q   R   A   P   L   Q   N   P   R   P   G   D   Y   L   L   N   K   T   A   S   T   E   G   I   A   S   K   K   S   L   E   L   K   K   R
3601/1201                               3631/1211                               3661/1221                               3691/1231
TAT CTG CTG GGT GAG CCG GCC AAT GGC GAT AAG ATC CAG AAG TCC GGA TCC ACT TCA GTG CTG GAT TCA CGC ATT CGC AGC TTC CAG TCG AAC ATA TCG GAG TGC CAG AAG CTT TTG AAT
Y   L   L   G   E   P   A   N   G   D   K   I   Q   K   S   G   S   T   S   V   L   D   S   R   I   R   S   F   Q   S   N   I   S   E   C   Q   K   L   L   N
3721/1241                               3751/1251                               3781/1261                               3811/1271
CCC AGC AGC GAC ATA AGT GCC GGC ATG CGA ACC TTC CTC GAT CGC ACA AAG TTG GGC GAA GGC AGC CAG ACG ACA CCC GGA CAG ACG AAC GAA CTA ATC CGT TCC GCC ACC AGC AAT GTG
P   S   S   D   I   S   A   G   M   R   T   F   L   D   R   T   K   L   G   E   G   S   Q   T   T   P   G   Q   T   N   E   L   I   R   S   A   T   S   N   V
3841/1281                               3871/1291                               3901/1301                               3931/1311
ATT AAC GAT CTG CGC GTG GAG CTT CGG ATA CAG AAA ACT GAC TCC AGC CAC TCC ACG GAC AAC GAG AAG GAA AAC GTT TTC GTG AAC TGT AAG AAC GAG CTG AAC AAG GGG ATG GAA TAC
I   N   D   L   R   V   E   L   R   I   Q   K   T   D   S   S   H   S   T   D   N   E   K   E   N   V   F   V   N   C   K   N   E   L   N   K   G   M   E   Y
3961/1321                               3991/1331                               4021/1341                               4051/1351
ACG GAT GCG GTC AAT GCC ACG CTG CTG GAC CAG CTG GCC AGA AAA AGT TCA CCC ACC ACG CCG ACG AAT AAG ACG GTG GTC GAG GTT ATT GAC CTG GTT ACA CCT GAG AAG CCA ATT GAC
T   D   A   V   N   A   T   L   L   D   Q   L   A   R   K   S   S   P   T   T   P   T   N   K   T   V   V   E   V   I   D   L   V   T   P   E   K   P   I   D
4081/1361                               4111/1371                               4141/1381                               4171/1391
ATT ATC GAT CTA ACG GCA CTG GAA ACG CCG AAA AAG CAG TTG GTC GAT GGT AGC GCC ATG GAT GTA GAT GAA CGC CTC ACA CCC GAT AGC AAC AAA ATC AGC GAA CTG CAG CAG GAA GTG
I   I   D   L   T   A   L   E   T   P   K   K   Q   L   V   D   G   S   A   M   D   V   D   E   R   L   T   P   D   S   N   K   I   S   E   L   Q   Q   E   V
4201/1401                               4231/1411                               4261/1421                               4291/1431
AAG GAG GAA CCC AAG CCG GAT GTC TCT AGG GAT GTG AAA GAA TGC ATA CCA GAT ATA CTG GGA CAC ATT AAG GAG GGA ACG GGA TCG AAG GAG CCA GGT GGA GAG GAC CAA CAG AGC CTG
K   E   E   P   K   P   D   V   S   R   D   V   K   E   C   I   P   D   I   L   G   H   I   K   E   G   T   G   S   K   E   P   G   G   E   D   Q   Q   S   L
4321/1441                               4351/1451                               4381/1461                               4411/1471
CTG GAG CAG TCG GAC GAA GAG AAG CGC GAC TCA CCG GAA AAG GAT GTG GCC GAA CAT GAG CTT TAT GAA CCG GAC AGT GTG CAG ATC CAG GTG CCC AAT ATC CCA TGG GAA AAA AGC AAG
L   E   Q   S   D   E   E   K   R   D   S   P   E   K   D   V   A   E   H   E   L   Y   E   P   D   S   V   Q   I   Q   V   P   N   I   P   W   E   K   S   K
```

FIGURE 16-2

```
                                                         DMICALshort.txt
4441/1481                               4471/1491                               4501/1501                               4531/1511
CCG GAG GTC ATG TCT ACC ACC GGC AGC AGT GGC TCC ATC TGC TCA AGC TCA GAC TCT TCT AGT ATT GAA GAC ATC CAG CAC TAC ATT TTG GAG TCC ACA ACT AGT CCA GAT ACT CAG ACA
P   E   V   M   S   T   T   G   S   S   G   S   I   C   S   S   S   D   S   S   S   I   E   D   I   Q   H   Y   I   L   E   S   T   T   S   P   D   T   Q   T
4561/1521                               4591/1531                               4621/1541                               4651/1551
GTT GGC GGA AAG CAC AAT GTG CCC CGT TTG GAG GTG CAC GAC ACA AGT GGT GCC CTG ATG CAG GTG GAC AGC CTG ATG ATT GTG AAC GGA AAG TAT ATT GGG GAT CCC GAG GAT GTC AAG
V   G   G   K   H   N   V   P   R   L   E   V   H   D   T   S   G   A   L   M   Q   V   D   S   L   M   I   V   N   G   K   Y   I   G   D   P   E   D   V   K
4681/1561                               4711/1571                               4741/1581                               4771/1591
TTC TTG GAT ATG CCG GCC AAT GTT ATT GTT CCG CCA GCA CCG GCG CTT AAA ACG AAT GAG CTG GAT ATG GAG GAT GAC CAA GAG GCG GAG GCG GAA CCA GTA ACT GCT ACT CCG GAG CCG
F   L   D   M   P   A   N   V   I   V   P   P   A   P   A   L   K   T   N   E   L   D   M   E   D   D   Q   E   A   E   A   E   P   V   T   A   T   P   E   P
4801/1601                               4831/1611                               4861/1621                               4891/1631
GTG GAA TGT ACG GTC ATC GAG GCT GAG CGC CGT GTT ACT GCT CCC CCT CCT TTG CCA GAG ATG GGT CCA CCC AAA CTG AAG TTC GAT AGC AAA AAT GAG AAC AAG ATC GAG AGC TTG AAG
V   E   C   T   V   I   E   A   E   R   R   V   T   A   P   P   P   L   P   E   M   G   P   P   K   L   K   F   D   S   K   N   E   N   K   I   E   S   L   K
4921/1641                               4951/1651                               4981/1661                               5011/1671
AAT CTT CCG TTG ATC GTA GAG AGC AAT GTG GAG CAC AGT CAG GCA GTG AAA CCC ATT ACT CTT AAC TTA AGC AAT CTG GCC AGG ACG CCG GAT ACA CCA ACC ACG CCC ACG GCG CAC GAT
N   L   P   L   I   V   E   S   N   V   E   H   S   Q   A   V   K   P   I   T   L   N   L   S   N   L   A   R   T   P   D   T   P   T   T   P   T   A   H   D
5041/1681                               5071/1691                               5101/1701                               5131/1711
AGC GAT AAA ACA CCC ACT GGG GAA ATT CTC TCG CGA GGA TCT GAC TCA GAA ACC GAG CAC ACT GGC ACT GGT CAG GTA CTA ACG GAG ACG GAA CTC TCC GAC TGG ACG GCC GAC GAC TGT
S   D   K   T   P   T   G   E   I   L   S   R   G   S   D   S   E   T   E   H   T   G   T   G   Q   V   L   T   E   T   E   L   S   D   W   T   A   D   D   C
5161/1721                               5191/1731                               5221/1741                               5251/1751
ATC TCG GAG AAC TTT GTT GAC TTG GAG TTC GCG CTT AAC TCT AAC AAG GGT ACG ATA AAA CGG CGC AAG GAT CGA CGA CGC AGT GGA GCA AGC AAA CTT CCC AGT GGC AAC GAG GTA ATC
I   S   E   N   F   V   D   L   E   F   A   L   N   S   N   K   G   T   I   K   R   R   K   D   R   R   R   S   G   A   S   K   L   P   S   G   N   E   V   I
5281/1761                               5311/1771                               5341/1781                               5371/1791
CAC GAG CTG GCC AGG CAG GCG CCA GTG GTG CAA ATG GAT GGA ATT CTT AGT GCC ATC GAC ATT GAT GAC ATA GAG TTC ATG GAC ACG GGT TCG GAG GGT TCT TGT GCT GAA GCT TAT CCC
H   E   L   A   R   Q   A   P   V   V   Q   M   D   G   I   L   S   A   I   D   I   D   D   I   E   F   M   D   T   G   S   E   G   S   C   A   E   A   Y   P
5401/1801                               5431/1811                               5461/1821                               5491/1831
GCA ACA AAT ACA GCT CTC ATT CAG AAT AGA GGT TAC ATG GAG TAC ATC GAG GCG GAG CCG AAA AAG ACG ACC CGC AAG GCA GCT CCA CCA TCC AGT TAC CCA GGA AAT TTA CCG CCT TTA
A   T   N   T   A   L   I   Q   N   R   G   Y   M   E   Y   I   E   A   E   P   K   K   T   T   R   K   A   A   P   P   S   S   Y   P   G   N   L   P   P   L
5521/1841                               5551/1851                               5581/1861                               5611/1871
ATG ACG AAG CGG GAC GAG AAA CTG GGC GTT GAT TAC ATT GAG CAG GGG GCG TAC ATA ATG CAC GAT GAT GCA AAG ACG CCT GTG AAT GAG GTG GCT CCT GCC ATG ACC CAG TCG CTA ACT
M   T   K   R   D   E   K   L   G   V   D   Y   I   E   Q   G   A   Y   I   M   H   D   D   A   K   T   P   V   N   E   V   A   P   A   M   T   Q   S   L   T
5641/1881                               5671/1891                               5701/1901                               5731/1911
GAC TCA ATC ACG CTC AAT GAA CTG GAT GAT GAC AGT ATG ATA ATA TCC CAA ACC CAG CCA ACG ACA ACG GAG GAA AGT GAG GCA CTG ACG GTG GTC ACC AGT CCA CTT GAC ACG TCC TCG
D   S   I   T   L   N   E   L   D   D   D   S   M   I   I   S   Q   T   Q   P   T   T   T   E   E   S   E   A   L   T   V   V   T   S   P   L   D   T   S   S
5761/1921                               5791/1931                               5821/1941                               5851/1951
CCC AGG GTT CTC GAT CAA TTT GCA TCC ATG TTG GCG GCG GGA AAA GGT GAC TCC ACA CCC AGT AGC TCA GAG CAA CAA CCA AAG ACG TCT ACG GTG ACG AGC AGC AGC ACT GGG CCC AAC
P   R   V   L   D   Q   F   A   S   M   L   A   A   G   K   G   D   S   T   P   S   S   S   E   Q   Q   P   K   T   S   T   V   T   S   S   S   T   G   P   N
5881/1961                               5911/1971                               5941/1981                               5971/1991
TCC TCG ACA ACA GGA AAC GTC TCG AAG GAG CCG CAG GAG GAG GAC CTG CAA ATC CAG TTT GAG TAT GTT CGA GCA CTG CAG CAG CGG ATA TCG CAG ATC AGC ACC CAA CGG CGT AAG AGC
S   S   T   T   G   N   V   S   K   E   P   Q   E   E   D   L   Q   I   Q   F   E   Y   V   R   A   L   Q   Q   R   I   S   Q   I   S   T   Q   R   R   K   S
6001/2001                               6031/2011                               6061/2021                               6091/2031
TCT AAG GGA GAG GCA CCT AAC CTG CAG CTA AAC AGT AGC GCA CCT GTG ATA GAA TCA GCC GAG GAT CCG GCC AAG CCC GCA GAG GAG CCT CTG GTC TCA ATG CGA CCG CGG ACC ACC AGC
S   K   G   E   A   P   N   L   Q   L   N   S   S   A   P   V   I   E   S   A   E   D   P   A   K   P   A   E   E   P   L   V   S   M   R   P   R   T   T   S
6121/2041                               6151/2051                               6181/2061                               6211/2071
ATT TCC GGA AAG GTA CCG GAG ATA CCC ACA CTT AGC AGC AAG CTG GAA GAG ATA ACC AAA GAA CGC ACT AAG CAA AAG GAT CTG ATT CAC GAC CTA GTC ATG GAC AAG TTG CAG TCG AAG
I   S   G   K   V   P   E   I   P   T   L   S   S   K   L   E   E   I   T   K   E   R   T   K   Q   K   D   L   I   H   D   L   V   M   D   K   L   Q   S   K
6241/2081                               6271/2091                               6301/2101                               6331/2111
AAG CAG CTA AAC GCT GAG AAG CGT CTG CAC CGG AGT CGA CAG CGC AGT TTG CTG ACC AGT GGC TAT GCC AGT GGC TCC AGC CTT AGT CCG ACG CCC AAG CTG GCT GCT GCT TGC AGT CCG
K   Q   L   N   A   E   K   R   L   H   R   S   R   Q   R   S   L   L   T   S   G   Y   A   S   G   S   S   L   S   P   T   P   K   L   A   A   A   C   S   P
6361/2121                               6391/2131                               6421/2141                               6451/2151
CAG GAT TCC AAC TGC TCT AGC CAA GCG CAC TAC CAC GCC TCC ACG GCG GAG GAG GCC CCG AAG CCG CCG GCG GAA AGG CCG TTG CAG AAG TCC GCC ACG TCC ACC TAT GTG TCG CCT TAT
Q   D   S   N   C   S   S   Q   A   H   Y   H   A   S   T   A   E   E   A   P   K   P   P   A   E   R   P   L   Q   K   S   A   T   S   T   Y   V   S   P   Y
6481/2161                               6511/2171                               6541/2181                               6571/2191
CGC ACT GTC CAA GCG CCC ACA CGT AGT GCT GAT CTC TAT AAG CCG CGC CCC TTC AGC GAA CAC ATC GAT TCG AAC GCT CTG GCG GGT TAC AAG CTC GGC AAG ACG GCC TCG TTT AAT GGC
R   T   V   Q   A   P   T   R   S   A   D   L   Y   K   P   R   P   F   S   E   H   I   D   S   N   A   L   A   G   Y   K   L   G   K   T   A   S   F   N   G
6601/2201                               6631/2211                               6661/2221                               6691/2231
GGC AAG TTG GGC GAC TTT GCG AAA CCC ATT GCC CCG GCG AGA GTT AAC CGA GGA GGA GGT GTC GCG ACC GCG GAT ATA GCC AAT ATT TCC GCG TCG ACG GAG AAC CTA AGA AGC GAG GCC
```

FIGURE 16-3

DMICALshort.txt

G K L G D F A K P I A P A R V N R G G G V A T A D I A N I S A S T E N L R S E A
6721/2241 6751/2251 6781/2261 6811/2271
AGG GCC AGG GCT CGT CTT AAG TCT AAC ACA GAG CTG GGC CTT AGT CCC GAG GAA AAG ATG CAG CTA ATA CGT TCA AGA TTG CAC TAC GAC CAA AAC AGA TCT CTG AAG CCG AAG CAA CTG
R A R A R L K S N T E L G L S P E E K M Q L I R S R L H Y D Q N R S L K P K Q L
6841/2281 6871/2291 6901/2301 6931/2311
GAG GAG ATG CCA TCC GGG GAT CTG GCG GCA CGT GCC CGC AAA ATG AGT GCC TCG AAG AGC GTC AAT GAT CTG GCC TAC ATG GTG GGA CAG CAG CAG CAG CAG CAG GTT GAG AAG GAT GCC
E E M P S G D L A A R A R K M S A S K S V N D L A Y M V G Q Q Q Q Q Q V E K D A
6961/2321 6991/2331 7021/2341 7051/2351
GTG CTC CAA GCC AAG GCG GCT GAC TTT ACA TCC GAT CCC AAT TTG GCG TCC GGT GGT CAG GAG AAG GCA GGC AAA ACT AAG TCC GGA CGC AGG CCA AAG GAT CCG GAG CGG CGT AAG AGT
V L Q A K A A D F T S D P N L A S G G Q E K A G K T K S G R R P K D P E R R K S
7081/2361 7111/2371 7141/2381 7171/2391
CTC ATA CAG TCG TTG TCC AGC TTC TTC CAA AAG GGA TCT GGA TCC GCG GCC TCC AGT TCC AAG GAG CAG GGC GGC GCT GTG GCT GCC GTC CAC TCT GAA CAG TCA GAG CGA CCA GGC ACC
L I Q S L S S F F Q K G S G S A A S S S K E Q G G A V A A V H S E Q S E R P G T
7201/2401 7231/2411 7261/2421 7291/2431
AGC AGC AGC GGC ACG CCC ACA ATA TCG GAT GCG GCG GGT GGA GGC GGA GGA GGA GGT GGC GTC TTC AGC AGA TTC CGC ATC TCG CCC AAG TCC AAG GAG AAG TCA AAG TCT TGC TTT GAT
S S S G T P T I S D A A G G G G G G G G V F S R F R I S P K S K E K S K S C F D
7321/2441 7351/2451 7381/2461 7411/2471
CTG AGG AAT TTC GGT TTT GGT GAC AAG GAT ATG CTG GTC TGC AAT GCA GCA TCT CCA GCA GGA GCC ACA TCC GCA TCA CAG AAA AAT CAC TCG CAA GAG TAT CTG AAC ACC ACG AAC AAC
L R N F G F G D K D M L V C N A A S P A G A T S A S Q K N H S Q E Y L N T T N N
7441/2481 7471/2491 7501/2501 7531/2511
AGT CGC TAT CGA AAG CAA ACG AAC ACT GCG AAA CCG AAA CCC GAA TCG TTC TCT TCA TCC AGT CCG CAG CTC TAT ATA CAC AAG CCC CAC CAC CTG GCC GCA GCT CAT CCC AGT GCC CTG
S R Y R K Q T N T A K P K P E S F S S S S P Q L Y I H K P H H L A A A H P S A L
7561/2521 7591/2531 7621/2541 7651/2551
GAC GAC CAG ACA CCA CCA CCC ATA CCG CCT CTT CCA CTG AAT TAT CAG AGA TCC GAT GAT GAG AGC TAC GCT AAC GAG ACA CGA GAG CAT AAG AAG CAA CGT GCC ATA TCG AAG GCT TCA
D D Q T P P P I P P L P L N Y Q R S D D E S Y A N E T R E H K K Q R A I S K A S
7681/2561 7711/2571 7741/2581 7771/2591
CGA CAA GCT GAG CTC AAG CGA TTG CGA ATC GCT CAA GAG ATT CAG CGG GAA CAG GAG GAG ATC GAG GTG CAA CTG AAG GAT CTG GAG GCA CGC GGC GTG CTT ATT GAG AAG GCC TTG CGA
R Q A E L K R L R I A Q E I Q R E Q E E I E V Q L K D L E A R G V L I E K A L R
7801/2601 7831/2611 7861/2621 7891/2631
GGC GAG GCG CAG AAT ATT GAA AAC CTG GAT GCG ACA AAG GAC AAC GAC GAG AAG CTA CTT AAG GAA CTT TTG GAG ATT TGG CGC AAC ATC ACA GCA CTC AAG AAA CGC GAT GAG GAA CTG
G E A Q N I E N L D A T K D N D E K L L K E L L E I W R N I T A L K K R D E E L
7921/2641 7951/2651 7981/2661 8011/2671
ACT ATA AGG CAA CAG GAA CTG CAA CTG GAG TAT CGG CAT GCC CAG CTG AAG GAA GAG CTC AAT CTG CGC TTG TCC TGC AAC AAA CTG GAC AAA AGC TCT GCC GAT GTG GCC GCC GAG GGA
T I R Q Q E L Q L E Y R H A Q L K E E L N L R L S C N K L D K S S A D V A A E G
8041/2681 8071/2691 8101/2701 8131/2711
GCA ATT CTC AAC GAG ATG CTG GAA ATT GTC GCC AAG CGA GCC GCC CTA CGA CCC ACA GCC TCC CAG CTC GAC CTC ACG GCA GCG GGA TCA GCA TCC ACG TCC GCC GAG GCA ACG GGC ATT
A I L N E M L E I V A K R A A L R P T A S Q L D L T A A G S A S T S A E A T G I
8161/2721 8191/2731
AAG CTG ACG GGA CAA CCG CAT GAC CAC GAA GAA TCG ATC ATT TGA
K L T G Q P H D H E E S I I *
ATAGCCAACTTTCCCTGCAGATTGCTGGTCTGCGAAGTGGTGCATGGACTGTTGAGCCTAAATACTTAAGATGAAACAAGGGAACAGGGGAGAGCGTACCAAAAAGGCAACTGAATATACAGAATTAACAATACTAAAACCCGTGAGCTGTAACCGAAAC
GAGCAACTAACTATTCAAAGTGGATAAGCTGGCTAAGAACAATTTCAATATTTATATAAGTAGAGAGCCTAACTAATATGGGAGAAGCGCGCAAGGTTAATGTAAAAACAGGCAGAGATGAAATCAAAACCAAAAAACTAATTTCGCAATTTAAAAGTAG
GGATTACGAATGTTTTTGCGAATACCTTTAGTTTCGTATTCGACCTAATTGTACGTTTTGTTAGTCGTTGCCACTGAAATCTGTCTGATTTTTGTATGTCCTATGCAATATGTTGTGTTTTTATTTAATGTAATGTAGTGAGCGTGTGCAAAGGGTATGTGATTTT
GAGCAAATGTAAATGGTGTATTTGAGTTTTCTTAGGATATTTTCACTTAGCGATTGTTTGTCGATTTCTATGTAAACCGTTTGAGTGTATTATTATTTATTTGACAAAGCACCTTCCGCACATCCTCCCGATCAGCACACCCATGTATGTATTA
TGTCATTTGCCCCCAATCGAATTCAAAGTGAACACAAAGAGATGCACACAAATATTTTCTAAATGCGAGAGCCTACACATAGCGATGTAAGCGAGAAAGCAACCATTAACTGTATACTAACGAAATCATTTAAATTATTGCGTAACTTTATTTATAACCT
GCATATATAAAAACATATTTATCGAATTTATTATGTAACTTTTCGATTTATTTTGAATTGAATAAAGATTTTATGAAACCTAA

FIGURE 16-4

HMICAL-Like1.txt

```
ggccaagccggggccccgaagccagagccggagccgggcgggccgcggggtc
1/1                                     31/11                                   61/21                                   91/31
atg gct ggg ccg cgg ggc gcg ctg ctg gcc tgg tgc cgc cgc cag tgc gag ggc tac cgc ggc gtg gag atc cgc gac ctg agc agc tcc ttc cgg gac ggc ctg gcc ttc tgc gcc atc
M   A   G   P   R   G   A   L   L   A   W   C   R   R   Q   C   E   G   Y   R   G   V   E   I   R   D   L   S   S   S   F   R   D   G   L   A   F   C   A   I
121/41                                  151/51                                  181/61                                  211/71
ctg cac cgg cac cgg ccc gac ctg cta gat ttt gat tcg ctt tcc aag gac aat gtc ttc gag aat aac cgt ttg gcc ttt gaa gtg gct gag aag gag ctg ggg atc ccc gct ctc ctg
L   H   R   H   R   P   D   L   L   D   F   D   S   L   S   K   D   N   V   F   E   N   N   R   L   A   F   E   V   A   E   K   E   L   G   I   P   A   L   L
241/81                                  271/91                                  301/101                                 331/111
gac ccc aat gac atg gtc tcc atg agc gtc cct gac tgc ctc agc atc atg acc tat gtg tcc cag tat tac aac cac ttc tgc agt cct ggc caa gct ggt gtc tcg cca ccc aga aag
D   P   N   D   M   V   S   M   S   V   P   D   C   L   S   I   M   T   Y   V   S   Q   Y   Y   N   H   F   C   S   P   G   Q   A   G   V   S   P   P   R   K
361/121                                 391/131                                 421/141                                 451/151
ggc ctt gca ccc tgt tcc ccg ccg tct gta gca ccc act cca gtg gaa cca gaa gat gtg gct cag ggc gag gag ctc tcc tca ggc agc ctg tca gag cag ggc acc ggc cag acc ccc
G   L   A   P   C   S   P   P   S   V   A   P   T   P   V   E   P   E   D   V   A   Q   G   E   E   L   S   S   G   S   L   S   E   Q   G   T   G   Q   T   P
481/161                                 511/171                                 541/181                                 571/191
agc agc acg tgc gca gcc tgc cag cag cat gtg cac ttg gtg cag cgc tac ctg gct gac ggc agg ctg tac cat cgc cac tgc ttc cgg tgt cgg cgg tgc tcc agc acc ctg ctc cct
S   S   T   C   A   A   C   Q   Q   H   V   H   L   V   Q   R   Y   L   A   D   G   R   L   Y   H   R   H   C   F   R   C   R   R   C   S   S   T   L   L   P
601/201                                 631/211                                 661/221                                 691/231
ggg gct tat gag aat ggg cct gag gag ggc acc ttt gtg tgt gca gaa cac tgt gcc agg ctg ggc ccg ggg aca cgg tcg ggg acc agg cct ggg ccc ttc tca cag cca aag cag cag
G   A   Y   E   N   G   P   E   E   G   T   F   V   C   A   E   H   C   A   R   L   G   P   G   T   R   S   G   T   R   P   G   P   F   S   Q   P   K   Q   Q
721/241                                 751/251                                 781/261                                 811/271
cac cag cag caa ctc gca gaa gat gcc aag gat gtt cca gga ggc ggc ccc agc tcc agt gct cct gca ggg gct gag gcc gat gga ccc aag gcc agc cct gag gcc cgg ccg cag atc
H   Q   Q   Q   L   A   E   D   A   K   D   V   P   G   G   G   P   S   S   S   A   P   A   G   A   E   A   D   G   P   K   A   S   P   E   A   R   P   Q   I
841/281                                 871/291                                 901/301                                 931/311
cct acc aag ccc cgg gtt cct ggc aaa cta cag gag ctg gcc agc ccc cct gcg ggc cgc ccc acc cct gcc ccc agg aag gcc tct gag agc acc acc cca gca ccc ccc acg ccc cgg
P   T   K   P   R   V   P   G   K   L   Q   E   L   A   S   P   P   A   G   R   P   T   P   A   P   R   K   A   S   E   S   T   T   P   A   P   P   T   P   R
961/321                                 991/331                                 1021/341                                1051/351
ccc cgc tcc agt ctg cag cag gag aac ctg gtg gag cag gct ggc agc agc agc ctg gtg aac ggg aga ctg cac gaa ctg cct gtc ccc aag ccg agg ggg aca ccg aag ccg tcc gag
P   R   S   S   L   Q   Q   E   N   L   V   E   Q   A   G   S   S   S   L   V   N   G   R   L   H   E   L   P   V   P   K   P   R   G   T   P   K   P   S   E
1081/361                                1111/371                                1141/381                                1171/391
ggg aca cca gcc ccc agg aag gac ccc cca tgg atc acg ctg gtg cag gca gaa cca aag aag aag cca gcc cca ctt ccc cca agc agc agc ccg ggg cca cca agc cag gac agc agg
G   T   P   A   P   R   K   D   P   P   W   I   T   L   V   Q   A   E   P   K   K   K   P   A   P   L   P   P   S   S   S   P   G   P   P   S   Q   D   S   R
1201/401                                1231/411                                1261/421                                1291/431
cag gtg gag aat gga ggc acc gag gag gtg gcc cag ccg agc cca acg gcc agc ctg gag tcc aaa ccc tat aac ccc ttt gag gag gag gag gag gac aag gag gaa gag gct cca gct
Q   V   E   N   G   G   T   E   E   V   A   Q   P   S   P   T   A   S   L   E   S   K   P   Y   N   P   F   E   E   E   E   E   D   K   E   E   E   A   P   A
1321/441                                1351/451                                1381/461                                1411/471
gca ccc agc ctg gcc acc agc cct gcc ctg ggc cac ccg gag tcc aca ccc aag tcc ctg cac ccc tgg tac ggc atc acc cct acc agc agc ccc aag aca aag aag cgc cct gcc ccg
A   P   S   L   A   T   S   P   A   L   G   H   P   E   S   T   P   K   S   L   H   P   W   Y   G   I   T   P   T   S   S   P   K   T   K   K   R   P   A   P
1441/481                                1471/491                                1501/501                                1531/511
cgc gca ccc agc gcg tcc cca ctg gct ctc cac gcc tcc cgc ctc tcg cac tcg gag ccg ccc tcg gcc aca cca tcg cca gcg ctc agc gtg gag agc ctg tcg tct gag agc gcc agc
R   A   P   S   A   S   P   L   A   L   H   A   S   R   L   S   H   S   E   P   P   S   A   T   P   S   P   A   L   S   V   E   S   L   S   S   E   S   A   S
1561/521                                1591/531                                1621/541                                1651/551
cag act gca ggt gca gag ctt ctg gag ccg cca gct gtg ccc aag agc tcc tca gag cct gct gtc cat gcc cct ggt acc cct gga aac cct gtc agc ctc tct acc aac tcc tcc ctg
Q   T   A   G   A   E   L   L   E   P   P   A   V   P   K   S   S   S   E   P   A   V   H   A   P   G   T   P   G   N   P   V   S   L   S   T   N   S   S   L
1681/561                                1711/571                                1741/581                                1771/591
gcc tcc tct ggg gaa cta gtg gag cct aga gtg gaa caa atg cct caa gcc agc cct ggc ctt gcc ccc agg acc agg ggc agc tca ggt ccc cag cca gcc aag ccc tgc agt ggc gcc
A   S   S   G   E   L   V   E   P   R   V   E   Q   M   P   Q   A   S   P   G   L   A   P   R   T   R   G   S   S   G   P   Q   P   A   K   P   C   S   G   A
1801/601                                1831/611                                1861/621                                1891/631
acc cca acg cct ctc ttg ttg gtt gga gac agg agc ccg gtg cct tcc cct gga agc tcg tcc cca cag ctg cag gta aag tcc tcc tgc aag gag aat cct ttt aac cgg aag cca tca
T   P   T   P   L   L   L   V   G   D   R   S   P   V   P   S   P   G   S   S   S   P   Q   L   Q   V   K   S   S   C   K   E   N   P   F   N   R   K   P   S
1921/641                                1951/651                                1981/661                                2011/671
cct gca gcg tcc cca gcc aca aag aag gcc acc aag gga tcc aag cca gtg agg cca cct gcc cct gga cac ggc ttt cca ctc atc aaa cgc aag gtc cag gct gac cag tac atc cct
P   A   A   S   P   A   T   K   K   A   T   K   G   S   K   P   V   R   P   P   A   P   G   H   G   F   P   L   I   K   R   K   V   Q   A   D   Q   Y   I   P
2041/681                                2071/691                                2101/701                                2131/711
gag gag gac atc cat gga gag atg gat acc att gag cgc cgg ctg gat gcc ctg gag cac cgt ggg gtg ctg ctg gag gag aag ctg cgt ggc ggc ctg aat gag ggc cgt gag gat gac
E   E   D   I   H   G   E   M   D   T   I   E   R   R   L   D   A   L   E   H   R   G   V   L   L   E   E   K   L   R   G   G   L   N   E   G   R   E   D   D
```

FIGURE 17-1

```
                                                HMICAL-Like1.txt
2161/721                                 2191/731                                 2221/741                                 2251/751
atg ctg gtg gac tgg ttc aag ctc atc cac gag aag cac cta ctg gtg cgg cga gag tcc gag ctc atc tat gtc ttc aag cag cag aac ctg gag cag cgc cag gct gat gtc gag tat
M   L   V   D   W   F   K   L   I   H   E   K   H   L   L   V   R   R   E   S   E   L   I   Y   V   F   K   Q   Q   N   L   E   Q   R   Q   A   D   V   E   Y
2281/761                                 2311/771                                 2341/781                                 2371/791
gag ctc cgg tgc ctc ctc aat aag cca gaa aag gac tgg acg gag gag gac cgg gcc cgg gag aag gtg ctg atg cag gag ctt gtg acc ctc att gag cag cgc aac gct atc atc aac
E   L   R   C   L   L   N   K   P   E   K   D   W   T   E   E   D   R   A   R   E   K   V   L   M   Q   E   L   V   T   L   I   E   Q   R   N   A   I   I   N
2401/801                                 2431/811                                 2461/821                                 2491/831
tgc ctg gat gag gac cgg cag agg gag gaa gag gaa gac aag atg ttg gaa gcc atg atc aag aag aaa gag ttc cag agg gag gct gaa cct gag ggc aag aag aag ggg aag ttc aag
C   L   D   E   D   R   Q   R   E   E   E   E   D   K   M   L   E   A   M   I   K   K   K   E   F   Q   R   E   A   E   P   E   G   K   K   K   G   K   F   K
2521/841                                 2551/851                                 2581/861                                 2611/871
acc atg aag atg ttg aaa ctg cta gga aac aaa cgt gat gcc aag agc aag tcc ccc aga gac aag agc taa cag cac gag aag cca gtt ggg gac tgc ccc ctc ctg gag cag ctc ctg
T   M   K   M   L   K   L   L   G   N   K   R   D   A   K   S   K   S   P   R   D   K   S   *
ggc tgt gct ctg ttt gaa ggg ggc gcc ctg ctc ccc tca gat cag tca gga gga aga tga cta agg gga ggg atc ctc tgg gtg atg gcc tct tcc tcc tca ggg acc tct gac tgc tct
ggg cca aag aat ctc ttg ttt ctt ctc cga gcc cca ggc agc ggt gat tca gcc ctg ccc aac ctg att ctg atg act gcg gat gct gtg acg gac cca agg ggc aaa tag ggt ccc agg
gtc cag gga ggg gcg cct gct gag cac ttc cgc ccc tca ccc tgc cca gcc cct gcc atg agc tct ggg ctg ggt ctc cgc ctc cag ggt tct gct ctt cca ggc agg cca gca agt ggc
gct ggg cca cac tgg ctt ctt cct gcc cca tcc ctg gct ctg agt ctc tgt ctt cct gtc ctg tgc agg cgc cct tgg atc tca gtt tcc ctc act cag gaa ctc tgt ttc tga agt ctt
cag tta agt ttg agt tta tga ctg agt ggc ctg tac tgt cag acg tga atg ggc ctg acg ggc aaa tcc atc cct ctc tcc ctc aca gtt cca gga gcg gct tcc ctc gtc tcc cct tac
tcc aca ggg agc ctc cct tgc cag gac cag ggc tgc gac ggc cat gct ggg gca ggt gag tgc tct gtt agc tgc tcc cag tgc tgt ccc cag gct gca gtt ctg gtc cct ggt tgt cag
gta gga agg gtg cac ttg aag cag gtg ctc atc tcg gtt cct taa cgt tta tag tct gac ccc tca ctt agg ctt tcc tct gcc acc ccg gtc cag gga aga ggc tcg ctc ccg ccc atg
gtc atc act ggt ctg tct gct ctg ttg tct gtt ctt tcc ctg act ccc tcc cac cga agg cct gat ggc tac tca ccc ctc tgg gat ggc tat ggg aga gga gga gtg atg ggg acc gcc
acc ttt tct gca gga aat gtg ccc agc agc tct tgg tca aag cac tgt tgc tat aag cta tct ctg gga tgc ctc tag gcc ccc ttc cct cta cac acc tct ggg aaa aga tta cac tgt
att aac tct cga gga gtt tcc tca cca ata aac aga caa cct caa ctg cca gtg ccc tgc agc ctc ggg cca cag cgg cag cct tgt ttg cct tcc cac ctg cct ctg cca cac ctg gtg
gct gaa cat ctc tgg tcg ccc aga ggc cat gtt ggg gcc atc ctc caa gag gga tct ctg ccc tca ccg cct gcc act ggg cag gat ccc ttt cct ctg cag gga gag gtg gct cct cgg
cca tgc agc ccc tgg cag gct cct tct aaa cat gcc tgt tga cct gga gct ggc gcc acc aac tcc agg gcc ttt cca ggg cca gac agg taa cac gca tga acc cga gtg aca gct ctg
acg ggc tgt ttc ggt gtc agg aga caa agc tgg cag ggg cag ggg tga act gga ggc aag tca agt cac ctg tgg cct gtg ggg ctg aat gtg ggc ccg gtg ttg cca gat cct ttg tca
taa gaa gct aga aat cca gat ttt atg tgt gtg taa ttt gta aat gct gaa agc tag cct gaa ttt ttt ttt ttt ttt ttt gag aca gag tct cgc tct gtc gcc cag gct gga gtg cag
tgg cgc gat ctc agc tca ctg caa gct ccg cct cct ggg ttc acg cca tcc tcc tgc ctc ggc ctc ctg agc agc tgg gac tac agg cgc atg cta cga cgc ctg gct aat ttt ttg tat
ttt tag tag aga cgg ggt ttc acc gtg tta acc agg atg gtc tcg atc tcc tga cct tgt gat cca ccc acc ttg gcc tcc caa agt gct ggg att aca ggc gtg agc cac cac gcc cgg
cca cta gcc tga att tca atc aag ggt tgg ctg ata ctg tgt gtc cag ggt gga ctg gat ttg tcc tgg ggg gtt ctc tgg ttt gct gcc tcc tga cca cat gat ggg gcc ttc gag gtc
gag gac aac tgt tcc cat tag att gca ccc tct gcc ctc agg ttc ttg agg gtg tgt gga cac aga ggc ttt cca tgg gat gtc cct gag ccg gcc ctt gat tgg ggc ctc acc att tac
agg gcc gtt tta ttc tgc aaa ccg aaa ctt ggg tca tgt gac ctg atg gga tta tgg gac tcc ctc cag gtg ccc gag aca agg ttg ata ttt cca aaa tat ttt ggt gat tta gtg gga
caa gca aat gac aga ata ccg gag aag gca ggg atc gtg ggt gtc agg agc cag agg gga ggg gga cag atg tgc tgt gta cag gac aag gtg tca ggt gac tcc ttc cca gca ggg cct
cgc aga tgc aca agc acg gag ctg gtg ggt ttt gcc caa gaa agg tca cgc ggc aca tgc agg gat tgg aac tcc cag gcc agg gct cta ggt cgc tcc cac ctt ttc atg ttt ctt tct
gtg gcc atg ggt ata gtg gaa aga cat aaa gct aaa gcc aac ttt taa tcc tga atg cac tgc ttg cca ggt aaa tgc cct tgg ttg tgg tat ctt gtt gag act tag ttt tca cag agg
gat aat gaa ccg ttg cag agg ttt att gag atc att aac aga gtg gaa ttc agc acc cgc cac agc agc cag cca tgg g
```

FIGURE 17-2

HMICAL-Like2.txt

```
cggagggcggttccgccgcgcggcgggcgcgccgcccac
1/1                                     31/11                                   61/21                                   91/31
atg gcg gcc atc agg gcg ctg caa cag tgg tgc cgg cag cag tgc gag ggc tac cgc gac gtg aat atc tgc aac atg acc acg tcg ttc cgc gac ggc ctg gct ttc tgc gcc atc ctg
M   A   A   I   R   A   L   Q   Q   W   C   R   Q   Q   C   E   G   Y   R   D   V   N   I   C   N   M   T   T   S   F   R   D   G   L   A   F   C   A   I   L
121/41                                  151/51                                  181/61                                  211/71
cac cgc cac cgg ccc gac ctc ata aac ttc agt gct ctc aag aag gaa aat att tat gaa aac aat aaa ctg gcc ttc cgc gtg gcc gag gag cac ttg ggc atc cca gcc ttg ctg gat
H   R   H   R   P   D   L   I   N   F   S   A   L   K   K   E   N   I   Y   E   N   N   K   L   A   F   R   V   A   E   E   H   L   G   I   P   A   L   L   D
241/81                                  271/91                                  301/101                                 331/111
gcc gag gac atg gtg gcc ttg aag gtg cct gac cgg ctg agc atc ttg acc tac gtg tcc cag tat tac aac tac ttc cac ggc cgc tcc ccc att ggg ggc atg gca ggc gtg aag agg
A   E   D   M   V   A   L   K   V   P   D   R   L   S   I   L   T   Y   V   S   Q   Y   Y   N   Y   F   H   G   R   S   P   I   G   G   M   A   G   V   K   R
361/121                                 391/131                                 421/141                                 451/151
gcc tcg gag gac tct gag gag gag ccg tca ggg aag aag gct cca gtc cag gcg gcc aag ctg ccc tcg ccc gcc cca gcc cgg aag cct cca cta tct cca gcc cag aca aac cct gtg
A   S   E   D   S   E   E   E   P   S   G   K   K   A   P   V   Q   A   A   K   L   P   S   P   A   P   A   R   K   P   P   L   S   P   A   Q   T   N   P   V
481/161                                 511/171                                 541/181                                 571/191
gtc cag agg agg aat gag ggt gca ggg ggc ccg ccc ccc aag act gac cag gca ttg gcg ggc agc ttg gtc agc agc acc tgc ggg gtc tgc ggc aag cac gtg cac ctg gta cag cgg
V   Q   R   R   N   E   G   A   G   G   P   P   P   K   T   D   Q   A   L   A   G   S   L   V   S   S   T   C   G   V   C   G   K   H   V   H   L   V   Q   R
601/201                                 631/211                                 661/221                                 691/231
cac ctg gcc gac ggg agg ctt tac cac cgg agc tgc ttc agg tgt aag cag tgc tcc tgc acg ctg cac tcg ggg gcc tac aag gcc aca gga gag ccg ggc acc ttc gtc tgc acc agc
H   L   A   D   G   R   L   Y   H   R   S   C   F   R   C   K   Q   C   S   C   T   L   H   S   G   A   Y   K   A   T   G   E   P   G   T   F   V   C   T   S
721/241                                 751/251                                 781/261                                 811/271
cac ctc ccc gca gcc gcc tct gca agc ccc aag ttg acg ggt ctg gtc ccc cga cag cca ggg gcc atg ggt gtg gat tcc agg acc tcc tgt tcc cca cag aag gcc cag gag gca aac
H   L   P   A   A   A   S   A   S   P   K   L   T   G   L   V   P   R   Q   P   G   A   M   G   V   D   S   R   T   S   C   S   P   Q   K   A   Q   E   A   N
841/281                                 871/291                                 901/301                                 931/311
aag gcc aga ccg ttg gcc tgg gag cct gct gcg ggc aac tcg cct gcc agg gct tcc gtt cca gct gca ccc aac cct gca gcc acc agc gcc acg tcc gtc cac gtg agg agc cca gcc
K   A   R   P   L   A   W   E   P   A   A   G   N   S   P   A   R   A   S   V   P   A   A   P   N   P   A   A   T   S   A   T   S   V   H   V   R   S   P   A
961/321                                 991/331                                 1021/341                                1051/351
agg ccc tct gag agc cgc ctg gcc ccc act ccc acg gag ggg aaa gtc cgc cct cgt gtg acc aat agc tcc ccg atg ggc tgg tcg tca gct gcc ccg tgc aca gca gcg gct gcc tcc
R   P   S   E   S   R   L   A   P   T   P   T   E   G   K   V   R   P   R   V   T   N   S   S   P   M   G   W   S   S   A   A   P   C   T   A   A   A   A   S
1081/361                                1111/371                                1141/381                                1171/391
cat ccc gcc gtg ccc ccg agt gcc cca gac cct cgc ccg gcc aca ccc cag ggc ggg gga gcc ccc cga gtg gca gct cct caa acc aca ctc agt tca agc tcc aca tct gca gcc acg
H   P   A   V   P   P   S   A   P   D   P   R   P   A   T   P   Q   G   G   G   A   P   R   V   A   A   P   Q   T   T   L   S   S   S   S   T   S   A   A   T
1201/401                                1231/411                                1261/421                                1291/431
gtg gac ccc cca gcc tgg acc ccg tcc gcc tcc agg acc cag cag gcc cgg aat aag ttt ttc caa aca tca gca gtg ccc ccc ggc acc agc ctt tct ggc aga ggt ccc acc ccg tca
V   D   P   P   A   W   T   P   S   A   S   R   T   Q   Q   A   R   N   K   F   F   Q   T   S   A   V   P   P   G   T   S   L   S   G   R   G   P   T   P   S
1321/441                                1351/451                                1381/461                                1411/471
ctt gtt cta tcc aag gac agc agc aag gag cag gcg cgg aac ttc ctc aag cag gcc ctc tca gcg ctg gaa gag gct ggc gct ccg gcg cct ggc agg ccc tcc cca gcc act gcc gct
L   V   L   S   K   D   S   S   K   E   Q   A   R   N   F   L   K   Q   A   L   S   A   L   E   E   A   G   A   P   A   P   G   R   P   S   P   A   T   A   A
1441/481                                1471/491                                1501/501                                1531/511
gtt ccc agt tct cag ccc aaa act gaa gca cca caa gca agt ccc tta gcc aag ccg tta cag tcc tcg tct ccc cgg gtg ctt ggc ctc cct tcg agg atg gaa ccg cca gcc ccg ctg
V   P   S   S   Q   P   K   T   E   A   P   Q   A   S   P   L   A   K   P   L   Q   S   S   S   P   R   V   L   G   L   P   S   R   M   E   P   P   A   P   L
1561/521                                1591/531                                1621/541                                1651/551
agc acg agc agt acc tct cag gca tcc gcg ttg ccc ccg gca ggc agg agg aac ttg gct gaa tcc tca ggg gtc ggc agg gtg ggt gct ggc tcc agg ccg aag cca gag gcc ccg atg
S   T   S   S   T   S   Q   A   S   A   L   P   P   A   G   R   R   N   L   A   E   S   S   G   V   G   R   V   G   A   G   S   R   P   K   P   E   A   P   M
1681/561                                1711/571                                1741/581                                1771/591
gca aag ggt aaa agc acc acc tta acg cag gac atg agc acc agc ctc cag gaa ggc cag gag gac ggg ccg gca gga tgg aga gcg aat ctg aag ccc gtg gac agg agg agc cca gct
A   K   G   K   S   T   T   L   T   Q   D   M   S   T   S   L   Q   E   G   Q   E   D   G   P   A   G   W   R   A   N   L   K   P   V   D   R   R   S   P   A
1801/601                                1831/611                                1861/621                                1891/631
gag agg act ctg aag ccc aag gaa cca cgg gcc ctg gca gag ccg agg gcg ggg gag gcc ccc agg aag gtc tca ggc agc ttt gct ggg agt gtc cac atc acc ctg acc ccc gtg agg
E   R   T   L   K   P   K   E   P   R   A   L   A   E   P   R   A   G   E   A   P   R   K   V   S   G   S   F   A   G   S   V   H   I   T   L   T   P   V   R
1921/641                                1951/651                                1981/661                                2011/671
cct gac agg acc cca cgc cca gcc agc cca gga ccc agc ctc cca gcc agg tcc ccc tcc cca ccc cgc cgc agg aga ctg gcc gtc cct gcc agc ctc gac gtt tgt gac aac tgg ctt
P   D   R   T   P   R   P   A   S   P   G   P   S   L   P   A   R   S   P   S   P   P   R   R   R   R   L   A   V   P   A   S   L   D   V   C   D   N   W   L
2041/681                                2071/691                                2101/701                                2131/711
cgg ccg gag ccc cct ggc cag gaa gcc cga gtg cag agc tgg aag gag gag gag aag aaa cct cac ctt cag ggc aga cca ggg aga ccc ttg tcc ccg gcc aat gtc cct gct ctg cct
R   P   E   P   P   G   Q   E   A   R   V   Q   S   W   K   E   E   E   K   K   P   H   L   Q   G   R   P   G   R   P   L   S   P   A   N   V   P   A   L   P
2161/721                                2191/731                                2221/741                                2251/751
```

FIGURE 18-1

```
                                    HMICAL-Like2.txt
ggc gag acg gtg acc tcc cca gtc agg ctg cac ccc gac tac ctc tcc ccg gag gag ata cag agg cag ctg cag gac atc gag agg cgg ctg gac gcc ctg gag ctc cgc ggc gtg gag
G   E   T   V   T   S   P   V   R   L   H   P   D   Y   L   S   P   E   E   I   Q   R   Q   L   Q   D   I   E   R   R   L   D   A   L   E   L   R   G   V   E
2281/761                                    2311/771                                    2341/781                                    2371/791
ctg gag aag cga ctg cgg gcg gcc gag gga gat gac gct gag gat agc ctc atg gtg gac tgg ttc tgg ctc att cac gag aag cag ctt ctg ctg aga cag gag tca gag ctg atg tac
L   E   K   R   L   R   A   A   E   G   D   D   A   E   D   S   L   M   V   D   W   F   W   L   I   H   E   K   Q   L   L   L   R   Q   E   S   E   L   M   Y
2401/801                                    2431/811                                    2461/821                                    2491/831
aag tcc aag gcc cag cgt ctg gag gag cag cag ctg gac atc gag ggc gag ctg cgc cgg ctc atg gcc aag ccc gag gct ctg aag tca ctg cag gag cgg cgg cgg gag cag gag ctg
K   S   K   A   Q   R   L   E   E   Q   Q   L   D   I   E   G   E   L   R   R   L   M   A   K   P   E   A   L   K   S   L   Q   E   R   R   R   E   Q   E   L
2521/841                                    2551/851                                    2581/861                                    2611/871
ctg gag cag tac gtg agc acc gtg aac gac cgc agt gac atc gtg gac tcg ctg gac gag gac cgg ctc cgg gaa caa gag gag gat cag atg ctg cgg gac atg att gag aag ctg ggc
L   E   Q   Y   V   S   T   V   N   D   R   S   D   I   V   D   S   L   D   E   D   R   L   R   E   Q   E   E   D   Q   M   L   R   D   M   I   E   K   L   G
2641/881                                    2671/891                                    2701/901                                    2731/911
ctc cag agg aag aag tcc aag ttc cgc ttg tcc aag atc tgg tca cca aaa agc aaa agc agc ccc tcc cag tag tag cca gta ggg ccg tgg gct cgg ccc gga cct ggc atc cgg act
L   Q   R   K   K   S   K   F   R   L   S   K   I   W   S   P   K   S   K   S   S   P   S   Q   *
tgg act cgg ggc cat ggg ctt ggc ccg gac ccg gaa ccc gga ctt gta ctc ggg gcc gtg ggc tcg gcc cgg acc cgg cat tcg gac ttg gac tcg gga agg gcc tcc tgt ccc tac aag
ggg cat gtg gac agc agg gac ctg cgc tac cgt ctg tgg tct caa taa aga aac cga cca cat ggc ccc gga ctc cta cgt gcc tcc gtg ttt gcc tcg gtg ccc ctg ggc ggc ccc gca
cct gcc tgg ggt ctc gcc tgc ctc cca gct ctt gcc tgc agg acg gac gtg ctc tgt ggc ctc ggg cag gag ccg acc ctc tca ggg cct ggc cct cga agg tgt ttc ccc ccg ggc cgg
cct gcc gtg acg atc aca gcc ctg ccg tct cag gtc acg gtt ttg gat tct gcg cta acg agg tct ccc agg tgg gct gga aac act gat gaa gac gga tag aaa aaa cac cag ccc acc
gga tta tta ata aag ata atc aca cac acc cct t
```

FIGURE 18-2

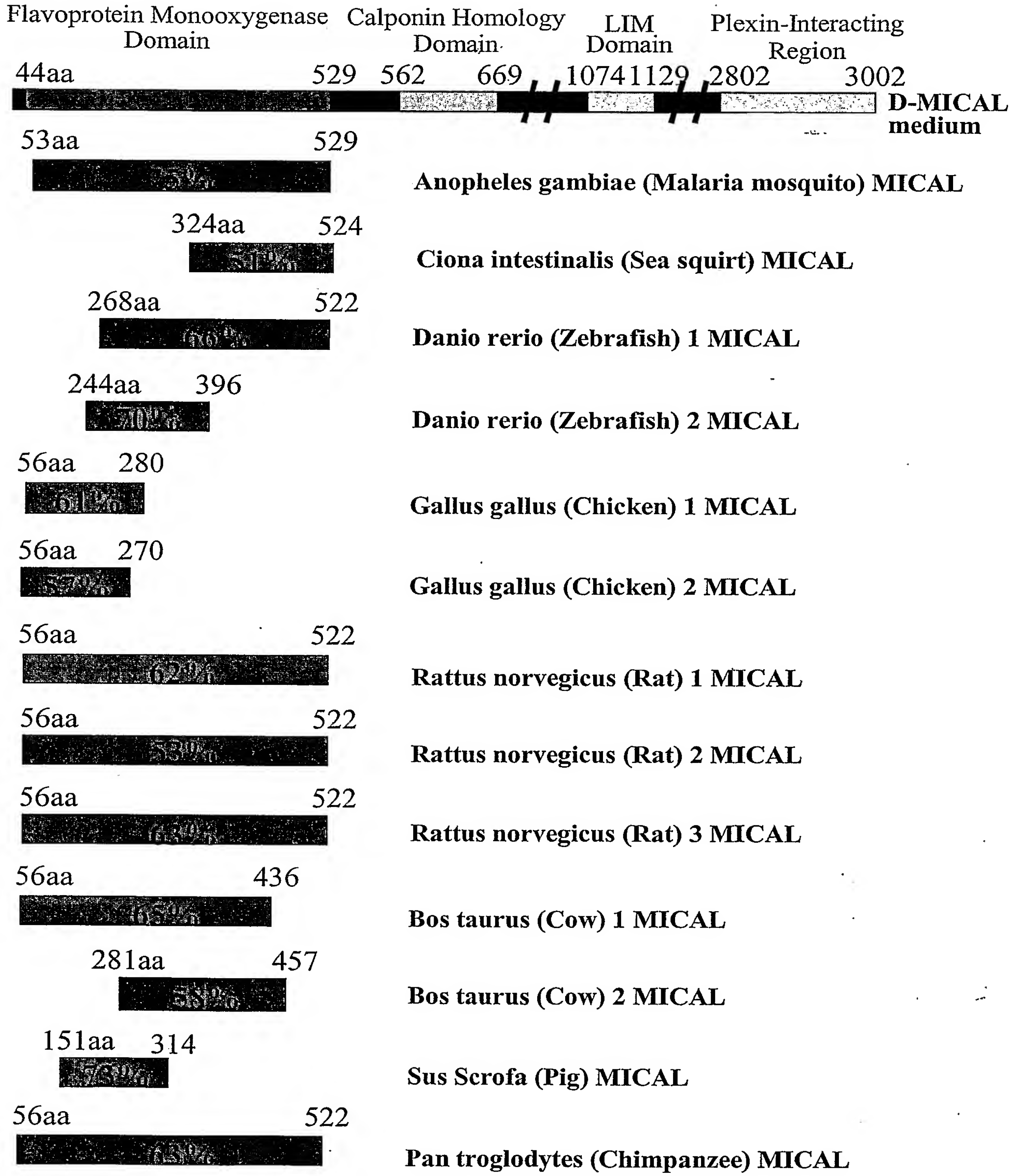
DMICAL-Like.txt

```
CACCGTTTTGCCTTGTTTACATTTAATGTAATTTGTACCAGATAAATCTTATTTGATTTTACTCAAGTCAAGTGGCACGATGTGCTTTTGTTGTTCACTTAGTTTACAGTTTGTGACTGCTATTTTCGTTTAGTTATATTCGCCAATAAGCCAGCAATT
AACCCACTAAA
1/1                                         31/11                                       61/21                                       91/31
ATG TCT GAT CGT CGT GGC ACA AAA GTT GGA ACT GGT ACG AAG GCT TTG GAG TAT TGG TGC CGA GTT GTG ACC CAA GGA TAT AAT GGG GTC AAG GTG GAG AAC ATG ACC ACT TCC TGG CGA
M   S   D   R   R   G   T   K   V   G   T   G   T   K   A   L   E   Y   W   C   R   V   V   T   Q   G   Y   N   G   V   K   V   E   N   M   T   T   S   W   R
121/41                                      151/51                                      181/61                                      211/71
AAT GGA CTT GCC TTC TGC GCC ATA ATA CAC CAC TTT CGG CCG GAT CTT ATA GAT TTC GAC CGA CTT AAA GCA GAC GAT ATC TAT GAG AAC AAC GAT TTG GCC TTT ACA ACG GCC GAG AAA
N   G   L   A   F   C   A   I   I   H   H   F   R   P   D   L   I   D   F   D   R   L   K   A   D   D   I   Y   E   N   N   D   L   A   F   T   T   A   E   K
241/81                                      271/91                                      301/101                                     331/111
TAT TTG GGA ATT CCC GCA CTG CTC GAT GCA GCT GAC ATG GTT TCG TAT GAA GTA CCT GAT AGA CTC TCC ATA CTA ACG TAT TTA TCC CAG TTT TAC AAG GTG CTC GGC AAG AGC CTG AAG
Y   L   G   I   P   A   L   L   D   A   A   D   M   V   S   Y   E   V   P   D   R   L   S   I   L   T   Y   L   S   Q   F   Y   K   V   L   G   K   S   L   K
361/121                                     391/131                                     421/141                                     451/151
CAC CCG AAG CCA GAG GAG CCA CTT GGC GAA GAG AGT GAG CCA CCG CAA AAG GTG ATG CAC ATT GTT GGG ATG CCT CGA CGT GAT AAG TGC CAG AAA TGC AAC CTT CCT GTT TTT CTC GCC
H   P   K   P   E   E   P   L   G   E   E   S   E   P   P   Q   K   V   M   H   I   V   G   M   P   R   R   D   K   C   Q   K   C   N   L   P   V   F   L   A
481/161                                     511/171                                     541/181                                     571/191
GAA AGG GTC CTT GTC GGC AAA CGG GCT TAT CAT CGC ACA TGC TTA AAA TGT GCA AGA TGT TCG TCA TTG CTC ACA CCC GGG AGT TTT TAT GAA ACG GAG GTC AAT AAC ATA TAC TGT TGC
E   R   V   L   V   G   K   R   A   Y   H   R   T   C   L   K   C   A   R   C   S   S   L   L   T   P   G   S   F   Y   E   T   E   V   N   N   I   Y   C   C
601/201                                     631/211                                     661/221                                     691/231
GAG ACT TGT CCC GAT GAA GAA AGT GAA CCG GAA TCT GAC ATT TTA AAG TTA AAG ACA ACC ACT ACT GAT TCT CCG AAT GAT AAA CAA ATG GTG GCA CAA AGT TCT GAT TAC TCT GAA GCT
E   T   C   P   D   E   E   S   E   P   E   S   D   I   L   K   L   K   T   T   T   T   D   S   P   N   D   K   Q   M   V   A   Q   S   S   D   Y   S   E   A
721/241                                     751/251                                     781/261                                     811/271
GAA GAT AAA CAA GAA GAC CTG GAA GAT AAT GAT ATA CGT ACT ACT GAT AAG CCT GAA AAC TTC CAA CCG CCG TCC AAC AAA GAT GAA CAA AAT AAT GAA CTA ACT ATT AAT CCG GTA AAC
E   D   K   Q   E   D   L   E   D   N   D   I   R   T   T   D   K   P   E   N   F   Q   P   P   S   N   K   D   E   Q   N   N   E   L   T   I   N   P   V   N
841/281                                     871/291                                     901/301                                     931/311
CCT ATA TTA TCA GAG GAG CGT AAA ATC TCT TTT ATT CCA CTA GAC GAA GAA GAT GGT GGC CTT ATT GAA CAA TAT AAT AAA TCA ACA ACT CCT GTA AAG CCC GCT ATA CCA GAG AAA CCA
P   I   L   S   E   E   R   K   I   S   F   I   P   L   D   E   E   D   G   G   L   I   E   Q   Y   N   K   S   T   T   P   V   K   P   A   I   P   E   K   P
961/321                                     991/331                                     1021/341                                    1051/351
AAA GTC TCT ACT CTT CCA CTA GAC GAT GAA CAA CAT GCT GGC GTT GAA CAA AAC AAT GAT CTG GCG GTT AGT CCG GAA AAT GAC ATT CCA AAA GAG AAA CTT AAG ATT TCT TCG GTC TCA
K   V   S   T   L   P   L   D   D   E   Q   H   A   G   V   E   Q   N   N   D   L   A   V   S   P   E   N   D   I   P   K   E   K   L   K   I   S   S   V   S
1081/361                                    1111/371                                    1141/381                                    1171/391
ATA TAT CTT GAA GAT GAC CGC CTT GTT GTA GAT GCA ATA CAT CCG GAT AAT TTA GAT AAA CAA GAA GCT TTG AAC AAT ACA TCA GAC GCA CTC ATT CCA GAA TCA CAG GAA GCA CCG ATT
I   Y   L   E   D   D   R   L   V   V   D   A   I   H   P   D   N   L   D   K   Q   E   A   L   N   N   T   S   D   A   L   I   P   E   S   Q   E   A   P   I
1201/401                                    1231/411                                    1261/421                                    1291/431
CCT GAA AAT AAC ACT CAA GTA GCC ATT AAA CCA GAA GAC CAT ATA AGC CCG CGT AAA GAA AAT AAA ATT TTC TCA AAC ACA GAA AGC TGC TCT AAG CAG GAA GGT GTT CTC CCA AAA CAA
P   E   N   N   T   Q   V   A   I   K   P   E   D   H   I   S   P   R   K   E   N   K   I   F   S   N   T   E   S   C   S   K   Q   E   G   V   L   P   K   Q
1321/441                                    1351/451                                    1381/461                                    1411/471
ATG GAT CTC GAG TCT CCC AAG GAC AAG GTA ATT GAG ACA AAA GCA TCT GAA ACT GAC TAT CCA GAG GAT CTA AAT CCC TTT AAA GAT GAT GAC TCA TCT AAA GGT GCC AAC CCA TTC GAC
M   D   L   E   S   P   K   D   K   V   I   E   T   K   A   S   E   T   D   Y   P   E   D   L   N   P   F   K   D   D   D   S   S   K   G   A   N   P   F   D
1441/481                                    1471/491                                    1501/501                                    1531/511
AGC TCT GAT GAT GAA GTG GAG CTT CTT AAA GCC ATT CCA GCC CAA CAA AGC AAA GGA AAA GTT GTA CCG CCC CGT CCA CCG CCA CCT AAA ATT GGC CTT TCA TCC ATC TCC AAT CCG TCA
S   S   D   D   E   V   E   L   L   K   A   I   P   A   Q   Q   S   K   G   K   V   V   P   P   R   P   P   P   P   K   I   G   L   S   S   I   S   N   P   S
1561/521                                    1591/531                                    1621/541                                    1651/551
GAG AAA CCA CAT TCC AGC CCT ACA CTT TCG CAT GGG AAA AAA ATG CCG ATG CCA ACA CCG AGA ATA TCT ATA TCA AAA ACC CAG ACT CCA GCA AAA CCA ATG ACG CAT CAA GGC CAG AAA
E   K   P   H   S   S   P   T   L   S   H   G   K   K   M   P   M   P   T   P   R   I   S   I   S   K   T   Q   T   P   A   K   P   M   T   H   Q   G   Q   K
1681/561                                    1711/571                                    1741/581                                    1771/591
AGT AGT ATT TCC TCT TCG TCT TCG GAG CAT TTG AAT AGC ATA AGA ACG TTC GAT CGC GGA GCA GAT GAC CGT GGC TCA AGC ATC TCC TTG CCA TCT GCA AAT GGG CCT CGC AAG CCA CTT
S   S   I   S   S   S   S   S   E   H   L   N   S   I   R   T   F   D   R   G   A   D   D   R   G   S   S   I   S   L   P   S   A   N   G   P   R   K   P   L
1801/601                                    1831/611                                    1861/621                                    1891/631
CGC GCC TCA GTG GGG AGT CCG CTG CGA AGC GAA GAA TCG AGT CCC ACT ACC AGT CTT AGC TCT ATT ACC TCT CCG ATG CGG AAG AAG CGC CAG GCA CCT TTG CCC CCA ATA CAA ACG GAC
R   A   S   V   G   S   P   L   R   S   E   E   S   S   P   T   T   S   L   S   S   I   T   S   P   M   R   K   K   R   Q   A   P   L   P   P   I   Q   T   D
1921/641                                    1951/651                                    1981/661                                    2011/671
TTT GAC AGT GAT CCT GGA TTT TCA AAA TTG TCC GAC GAA CAA AAG GCA TTG TTG CAC ACT CAG CTT AAG GCC CCA AAT CTG GGT GAT TCA ACC AGA AGA CTA ATA CCA CTT GAC CAG AGT
F   D   S   D   P   G   F   S   K   L   S   D   E   Q   K   A   L   L   H   T   Q   L   K   A   P   N   L   G   D   S   T   R   R   L   I   P   L   D   Q   S
2041/681                                    2071/691                                    2101/701                                    2131/711
CTG CTA TCC GAT GAG GCA ACC GAG TCG AGT AAT TAC GAC GAG AGC TTA AGT ACA TCC AAT GCC GAC GAG GAG GTA AAT GTG GTA TAC CGA CGC ATT TTG GTG CCG CCA ACT CAA CCC GAA
L   L   S   D   E   A   T   E   S   S   N   Y   D   E   S   L   S   T   S   N   A   D   E   E   V   N   V   V   Y   R   R   I   L   V   P   P   T   Q   P   E
```

FIGURE 19-1

```
                                                                              DMICAL-Like.txt
2161/721                                 2191/731                                 2221/741                                 2251/751
AAC ACT GTT GAA CGT AGC AAG GAG GAT CAA AAA TCG CCT ATC GTG TAT AAC GAC TTC GAT AGA AAC GTA AGC CCA TTG GGG CAC AAT AAA TCC ACT CAT GGG AAA TGG AAG AGG CGA AAG
N   T   V   E   R   S   K   E   D   Q   K   S   P   I   V   Y   N   D   F   D   R   N   V   S   P   L   G   H   N   K   S   T   H   G   K   W   K   R   R   K
2281/761                                 2311/771                                 2341/781                                 2371/791
GGA CCA GCA CCA GCG GTT CCA ATA CCA CCG CGC AAA GTC TTA CAA AGG CTT CCG CTG CAA GAA ATT CGC CAC GAG TTC GAA ATT ATT GCG GTA CAG CAG CTG GGT CTT GAG AAA CAA GGC
G   P   A   P   A   V   P   I   P   P   R   K   V   L   Q   R   L   P   L   Q   E   I   R   H   E   F   E   I   I   A   V   Q   Q   L   G   L   E   K   Q   G
2401/801                                 2431/811                                 2461/821                                 2491/831
GTT ATT CTG GAG AAA ATG ATC AGG GAT CGC TGT GAG CGT TCC TTA GAT GCC ACC GAT ACT GAT GGT CCC GAA TCC GCA GAA GTG CTA ACC AAC TCA AAG GAA GTG GAA GAT CTT ATA TTG
V   I   L   E   K   M   I   R   D   R   C   E   R   S   L   D   A   T   D   T   D   G   P   E   S   A   E   V   L   T   N   S   K   E   V   E   D   L   I   L
2521/841                                 2551/851                                 2581/861                                 2611/871
CAA CTG TTC GAG CTG GTC AAC GAG AAA AAC GAA TTG TTC CGC AGA CAA GCG GAA CTG ATG TAC CTT CGA CGT CAG CAT CGC CTG GAG CAG GAG CAG GCG GAC ATA GAG CAT GAG ATA CGA
Q   L   F   E   L   V   N   E   K   N   E   L   F   R   R   Q   A   E   L   M   Y   L   R   R   Q   H   R   L   E   Q   E   Q   A   D   I   E   H   E   I   R
2641/881                                 2671/891                                 2701/901                                 2731/911
GTA TTA ATG GGG CAA CCG GAG CAC AAC AAA ACC GAT TCG GAT AAA GCC CAT GAG GAA GTA TTA ATC AAT CGC CTT GTG AAG GTG GTT GAA ATG CGT AAC GAA GTA ATT GAT AGC CTA GAG
V   L   M   G   Q   P   E   H   N   K   T   D   S   D   K   A   H   E   E   V   L   I   N   R   L   V   K   V   V   E   M   R   N   E   V   I   D   S   L   E
2761/921                                 2791/931                                 2821/941                                 2851/951
ACT GAC CGA GTT CGC GAG GCG CGA GAG GAT ATG AGC ATC AAG AAT CGG CTT CAT ATA TAC AAT TCT GAG CGC GAG GAA CCA CCA GCA CAT CCG AGA AGC GCT GAC AAA TCA TCC AAA AAG
T   D   R   V   R   E   A   R   E   D   M   S   I   K   N   R   L   H   I   Y   N   S   E   R   E   E   P   P   A   H   P   R   S   A   D   K   S   S   K   K
2881/961                                 2911/971                                 2941/981                                 2971/991
CTG TCT AAA AAG GAA CGA AAA AAA CTT AAG GAG GAA AAT AAG CTG GGC AAG GGC AAA AAG TCG GAC CTC GAC AAG GAC GTC GAT GAG TCC GAA CAG GCA CCA GCT TTG GAA AAG GTT AAA
L   S   K   K   E   R   K   K   L   K   E   E   N   K   L   G   K   G   K   K   S   D   L   D   K   D   V   D   E   S   E   Q   A   P   A   L   E   K   V   K
3001/1001                                3031/1011                                3061/1021                                3091/1031
AAG AAG CGA AAC TTG TTC TTT TTG AAA ATG TAG TTG GGG GCT GGT TGA GTT GAA GTG TTT ACG CAT CGA ACA ATT GAA ATT TAT TTT ATA TAA ATT TGT AAC CTA TTA AGT TTC TAT TGC
K   K   R   N   L   F   F   L   K   M   *
CAT ATG TGA GCA TTT CAA TAA ATA AAG TAT ATA TTT TCC TTT AAA AAA AAA AAA AAA AAA A
```

FIGURE 19-2



Flavoprotein Monooxygenase Domain
Calponin Homology Domain
LIM Domain
Plexin-Interacting Region
44aa
529
562
669
1074
1129
2802
3002
D-MICAL medium
53aa
529
Anopheles gambiae (Malaria mosquito) MICAL
324aa
524
Ciona intestinalis (Sea squirt) MICAL
268aa
522
Danio rerio (Zebrafish) 1 MICAL
244aa
396
Danio rerio (Zebrafish) 2 MICAL
56aa
280
Gallus gallus (Chicken) 1 MICAL
56aa
270
Gallus gallus (Chicken) 2 MICAL
56aa
522
Rattus norvegicus (Rat) 1 MICAL
56aa
522
Rattus norvegicus (Rat) 2 MICAL
56aa
522
Rattus norvegicus (Rat) 3 MICAL
56aa
436
Bos taurus (Cow) 1 MICAL
281aa
457
Bos taurus (Cow) 2 MICAL
151aa
314
Sus Scrofa (Pig) MICAL
56aa
522
Pan troglodytes (Chimpanzee) MICAL

FIGURE 20

SEQUENCE LISTING

<110> THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
KOLODKIN, Alex L.
TERMAN, Jon R.
MAO, Tianyi
PASTERKAMP, Ronald J.
YU, Hung-Hsiang

<120> MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPTIDES, AND METHODS OF USING THE SAME

<130> JHU1840-3WO

<150> US 60/388,325
<151> 2002-06-13

<150> US 60/384,302
<151> 2002-05-30

<150> US 60/354,178
<151> 2002-02-04

<160> 40

<170> PatentIn version 3.1

<210> 1
<211> 3385
<212> DNA
<213> Homo sapiens

<400> 1
ccacctctcc agccactcat ctctgcccag ctgctgccct ccccaggagg cctccatggc 60

ttcacctacc tccaccaacc cagcgcatgc ccactttgag agcttcctgc aggcccagct 120

gtgccaggac gtgctgagca gcttccagga gctgtgtggg gccctggggc tggaacccgg 180

tggggggctg ccccagtacc acaagatcaa ggaccagctc aactactgga gcgccaagtc 240

actgtggacc aagctggaca agcgagcagg ccagcctgtc taccagcagg gccgggcctg 300

caccagcacc aagtgcctgg tggtgggtgc tggaccttgc gggctgcggg tcgctgtgga 360

gctggcgctg ctgggggccc gagtggtgct ggtggaaaag cgcaccaagt tctctcgcca 420

caacgtgctc cacctctggc ccttcaccat ccacgacctg cgggcactcg gtgctaagaa 480

gttctacggg cgcttctgca ccggcaccct ggaccacatc agcatcaggc agctccagct 540

gcttctgctg aaggtagcat tgctgctggg ggtggaaatt cactggggtg tcactttcac 600

tggcctccag ccccctccta ggaaggggag tggctggcgt gcccagctcc aacccaaccc 660

ccctgcccag ctggccaact atgaatttga cgtccttatc tcggctgcag gaggtaaatt 720

cgtccctgaa ggcttcaaag ttcgagaaat gcgaggcaaa ctggccattg gcatcacagc 780

caactttgtg aatggacgca ccgtggagga gacacaggtg ccggagatca gtggtgtagc 840

caggatctac aaccagagct tcttccagag ccttctcaaa gccacaggca ttgatctgga 900
gaacattgtg tactacaagg acgacaccca ctactttgtg atgacagcca agaagcagtg 960
cctgctgcgg ctgggggtgc tgcgccagga ctggccagac accaatcggc tgctgggcag 1020
tgccaatgtg gtgcccgagg ctctgcagcg ctttacccgg gcagctgctg actttgccac 1080
ccatggcaag ctcgggaaac tagagtttgc ccaggatgcc catgggcagc ctgatgtctc 1140
tgcctttgac ttcacgagca tgatgcgggc agagagttct gctcgtgtgc aagagaagca 1200
tggcgcccgc ctgctgctgg gactggtggg ggactgcctg gtggagccct tctggcccct 1260
gggcactgga gtggcacggg gcttcctggc agcctttgat gcagcctgga tggtgaagcg 1320
gtgggcagag ggcgctgagt ccctagaggt gttggctgag cgtgagagcc tgtaccagct 1380
tctgtcacag acatccccag aaaacatgca tcgcaatgtg gcccagtatg ggctggaccc 1440
agccacccgc taccccaacc tgaacctccg ggcagtgacc cccaatcagg tacgagacct 1500
gtatgatgtg ctagccaagg agcctgtgca gaggaacaac gacaagacag atacagggat 1560
gccagccacc gggtcggcag gcacccagga ggagctgcta cgctggtgcc aggagcagac 1620
agctgggtac ccgggagtcc acgtctccga tttgtcttcc tcctgggctg atgggctagc 1680
tctgtgtgcc ctggtgtacc ggctgcagcc tggcctgctg gaaccctcag agctgcaggg 1740
gctgggagct ctggaagcaa ctgcttgggc actaaaggtg gcagagaatg agctgggcat 1800
cacaccggtg gtgtctgcac aggccgtggt agcagggagt gacccactgg gcctcattgc 1860
ctacctcagc cacttccaca gtgccttcaa gagcatggcc cacagcccag gccctgtcag 1920
ccaggcctcc ccagggacct ccagtgctgt attattcctt agtaaacttc agaggaccct 1980
gcagcgatcc cgggccaagg aaaatgcaga ggatgctggt ggcaagaagc tgcgcttgga 2040
gatggaggcc gagaccccaa gtactgaggt gccacctgac ccagagcctg gtgtacccct 2100
gacaccccca tcccaacacc aggaggccgg tgctggggac ctgtgtgcac tttgtgggga 2160
acacctctat gtcctggaac gcctctgtgt caacggccat ttcttccacc ggagctgctt 2220
ccgctgccat acctgtgagg ccacactgtg gccaggtggc tacgagcagc acccaggaga 2280
tggacatttc tactgcctcc agcacctgcc ccagacagac cacaaagcgg aaggcagcga 2340
tagaggccct gagagtccgg agctccccac accaagtgag aatagcatgc caccaggcct 2400
ctcaactccc acagcctcgc aggagggggc cggtcctgtt ccagatccca gccagcccac 2460
ccgtcggcag atccgcctct ccagcccgga gcgccagcgg ttgtcctccc ttaaccttac 2520
ccctgacccg gaaatggagc ctccacccaa gcctccccgc agctgctccg ccttggcccg 2580
ccacgccctg gagagcagct ttgtgggctg ggggctgcca gtccagagcc ctcaagctct 2640
tgtggccatg gagaaggagg aaaaagagag tcccttctcc agtgaagagg aagaagaaga 2700

tgtgcctttg gactcagatg tggaacaggc cctgcagacc tttgccaaga cctcaggcac 2760

catgaataac tacccaacat ggcgtcggac tctgctgcgc cgtgcgaagg aggaggagat 2820

gaagaggttc tgcaaggccc agaccatcca acggcgacta aatgagattg aggctgcctt 2880

gagggagcta gaggccgagg gcgtgaagct ggagctggcc ttgaggcgcc agagcagttc 2940

cccagaacag caaaagaaac tatgggtagg acagctgcta cagctcgttg acaagaacaa 3000

cagcctggtg gctgaggagg ccgagctcat gatcacggtg caggaattga atctggagga 3060

gaaacagtgg cagctggacc aggagctacg aggctacatg aaccgggaag aaaacctaaa 3120

gacagctgct gatcggcagg ctgaggacca ggtcctgagg aagctggtgg atttggtcaa 3180

ccagagagat gccctcatcc gcttccagga ggagcgcagg ctcagcgagc tggccttggg 3240

gacaggggcc cagggctaga cgagggtggg ccgtctgctt tcgttcccac aaagaaagca 3300

cctcacccca gcacagtgcc acccctgttc atctgggctg cctggcagag agccttgctg 3360

tttacaatta aaatgtttct gccac 3385

<210> 2
<211> 863
<212> PRT
<213> Homo sapiens

<400> 2

Met Ala Gly Pro Arg Gly Ala Leu Leu Ala Trp Cys Arg Arg Gln Cys
1   5   10   15

Glu Gly Tyr Arg Gly Val Glu Ile Arg Asp Leu Ser Ser Ser Phe Arg
            20   25   30

Asp Gly Leu Ala Phe Cys Ala Ile Leu His Arg His Arg Pro Asp Leu
        35   40   45

Leu Asp Phe Asp Ser Leu Ser Lys Asp Asn Val Phe Glu Asn Asn Arg
    50   55   60

Leu Ala Phe Glu Val Ala Glu Lys Glu Leu Gly Ile Pro Ala Leu Leu
65   70   75   80

Asp Pro Asn Asp Met Val Ser Met Ser Val Pro Asp Cys Leu Ser Ile
                85   90   95

Met Thr Tyr Val Ser Gln Tyr Tyr Asn His Phe Cys Ser Pro Gly Gln
            100   105   110

Ala Gly Val Ser Pro Pro Arg Lys Gly Leu Ala Pro Cys Ser Pro Pro
115 120 125

Ser Val Ala Pro Thr Pro Val Glu Pro Glu Asp Val Ala Gln Gly Glu
130 135 140

Glu Leu Ser Ser Gly Ser Leu Ser Glu Gln Gly Thr Gly Gln Thr Pro
145 150 155 160

Ser Ser Thr Cys Ala Ala Cys Gln Gln His Val His Leu Val Gln Arg
165 170 175

Tyr Leu Ala Asp Gly Arg Leu Tyr His Arg His Cys Phe Arg Cys Arg
180 185 190

Arg Cys Ser Ser Thr Leu Leu Pro Gly Ala Tyr Glu Asn Gly Pro Glu
195 200 205

Glu Gly Thr Phe Val Cys Ala Glu His Cys Ala Arg Leu Gly Pro Gly
210 215 220

Thr Arg Ser Gly Thr Arg Pro Gly Pro Phe Ser Gln Pro Lys Gln Gln
225 230 235 240

His Gln Gln Gln Leu Ala Glu Asp Ala Lys Asp Val Pro Gly Gly Gly
245 250 255

Pro Ser Ser Ser Ala Pro Ala Gly Ala Glu Ala Asp Gly Pro Lys Ala
260 265 270

Ser Pro Glu Ala Arg Pro Gln Ile Pro Thr Lys Pro Arg Val Pro Gly
275 280 285

Lys Leu Gln Glu Leu Ala Ser Pro Pro Ala Gly Arg Pro Thr Pro Ala
290 295 300

Pro Arg Lys Ala Ser Glu Ser Thr Thr Pro Ala Pro Pro Thr Pro Arg
305 310 315 320

Pro Arg Ser Ser Leu Gln Gln Glu Asn Leu Val Glu Gln Ala Gly Ser
325 330 335

Ser Ser Leu Val Asn Gly Arg Leu His Glu Leu Pro Val Pro Lys Pro
340 345 350

Arg Gly Thr Pro Lys Pro Ser Glu Gly Thr Pro Ala Pro Arg Lys Asp

355 360 365

Pro Pro Trp Ile Thr Leu Val Gln Ala Glu Pro Lys Lys Lys Pro Ala
370 375 380

Pro Leu Pro Pro Ser Ser Ser Pro Gly Pro Pro Ser Gln Asp Ser Arg
385 390 395 400

Gln Val Glu Asn Gly Gly Thr Glu Glu Val Ala Gln Pro Ser Pro Thr
405 410 415

Ala Ser Leu Glu Ser Lys Pro Tyr Asn Pro Phe Glu Glu Glu Glu Glu
420 425 430

Asp Lys Glu Glu Glu Ala Pro Ala Ala Pro Ser Leu Ala Thr Ser Pro
435 440 445

Ala Leu Gly His Pro Glu Ser Thr Pro Lys Ser Leu His Pro Trp Tyr
450 455 460

Gly Ile Thr Pro Thr Ser Ser Pro Lys Thr Lys Lys Arg Pro Ala Pro
465 470 475 480

Arg Ala Pro Ser Ala Ser Pro Leu Ala Leu His Ala Ser Arg Leu Ser
485 490 495

His Ser Glu Pro Pro Ser Ala Thr Pro Ser Pro Ala Leu Ser Val Glu
500 505 510

Ser Leu Ser Ser Glu Ser Ala Ser Gln Thr Ala Gly Ala Glu Leu Leu
515 520 525

Glu Pro Pro Ala Val Pro Lys Ser Ser Ser Glu Pro Ala Val His Ala
530 535 540

Pro Gly Thr Pro Gly Asn Pro Val Ser Leu Ser Thr Asn Ser Ser Leu
545 550 555 560

Ala Ser Ser Gly Glu Leu Val Glu Pro Arg Val Glu Gln Met Pro Gln
565 570 575

Ala Ser Pro Gly Leu Ala Pro Arg Thr Arg Gly Ser Ser Gly Pro Gln
580 585 590

Pro Ala Lys Pro Cys Ser Gly Ala Thr Pro Thr Pro Leu Leu Leu Val
595 600 605

```
Gly Asp Arg Ser Pro Val Pro Ser Pro Gly Ser Ser Ser Pro Gln Leu
    610                 615                 620

Gln Val Lys Ser Ser Cys Lys Glu Asn Pro Phe Asn Arg Lys Pro Ser
625                 630                 635                 640

Pro Ala Ala Ser Pro Ala Thr Lys Lys Ala Thr Lys Gly Ser Lys Pro
                645                 650                 655

Val Arg Pro Pro Ala Pro Gly His Gly Phe Pro Leu Ile Lys Arg Lys
            660                 665                 670

Val Gln Ala Asp Gln Tyr Ile Pro Glu Glu Asp Ile His Gly Glu Met
        675                 680                 685

Asp Thr Ile Glu Arg Arg Leu Asp Ala Leu Glu His Arg Gly Val Leu
    690                 695                 700

Leu Glu Glu Lys Leu Arg Gly Gly Leu Asn Glu Gly Arg Glu Asp Asp
705                 710                 715                 720

Met Leu Val Asp Trp Phe Lys Leu Ile His Glu Lys His Leu Leu Val
                725                 730                 735

Arg Arg Glu Ser Glu Leu Ile Tyr Val Phe Lys Gln Gln Asn Leu Glu
            740                 745                 750

Gln Arg Gln Ala Asp Val Glu Tyr Glu Leu Arg Cys Leu Leu Asn Lys
        755                 760                 765

Pro Glu Lys Asp Trp Thr Glu Glu Asp Arg Ala Arg Glu Lys Val Leu
    770                 775                 780

Met Gln Glu Leu Val Thr Leu Ile Glu Gln Arg Asn Ala Ile Ile Asn
785                 790                 795                 800

Cys Leu Asp Glu Asp Arg Gln Arg Glu Glu Glu Glu Asp Lys Met Leu
                805                 810                 815

Glu Ala Met Ile Lys Lys Lys Glu Phe Gln Arg Glu Ala Glu Pro Glu
            820                 825                 830

Gly Lys Lys Lys Gly Lys Phe Lys Thr Met Lys Met Leu Lys Leu Leu
        835                 840                 845
```

Gly Asn Lys Arg Asp Ala Lys Ser Lys Ser Pro Arg Asp Lys Ser
850 855 860

<210> 3
<211> 6008
<212> DNA
<213> Homo sapiens

<400> 3
ccgggccgcc tcgctcgctc ccagctctgt cagtggcccg cggggcccga tcgctgcgcc 60

cgcggccagg gccgaggcag gcctgacccg gggccgggca gcccgcgcga ctttcggaac 120

atggcaaccc gtgtgtgtct catcccagaa agagaagact ttaaccactg tgatgcctga 180

gaatccagtg tgacgtttct ccagatactt catgctgttc acctgtgtcc tcgccgcacc 240

actgccgcac acgactcctg aaccatgggg gaaaacgagg atgagaagca ggcccaggcg 300

gggcaggttt ttgagaactt tgtccaggca tccacgtgca aaggtaccct ccaggccttc 360

aacattctca cacgacacct ggacctagac cctctggacc acagaaactt ttattccaag 420

ctcaagtcca aggtgaccac ctggaaagcc aaagccctgt ggtacaaatt ggataagcgt 480

ggttcccaca aagagtataa gcgagggaag tcgtgcacga acaccaagtg tctcatagtt 540

gggggaggac cctgtggctt gcgcactgcc attgaacttg cctacctggg agccaaagtg 600

gtcgtggtgg agaagaggga ctccttctcc cggaacaacg tgctacacct ctggcctttc 660

accatccatg accttcgggg cctgggagcc aagaagttct atgggaagtt ctgtgctggc 720

tccatcgacc atatcagtat tcgccaacta cagctcatcc tattcaaggt ggccctgatg 780

ctgggagttg aaatccatgt gaatgtggag ttcgtgaagg ttctagagcc tcctgaagat 840

caagaaaatc aaaaaattgg ctggcgggca gaatttctcc ctacagacca ttctctgtcg 900

gagtttgagt ttgacgtcat cattggtgcc gatggccgca ggaacaccct ggaagggttc 960

agaagaaaag aattccgtgg gaagctggcg attgccatca ccgccaactt cataaacaga 1020

aacagcacag cggaagccaa ggtggaagag attagtggtg tggctttcat cttcaatcag 1080

aaatttttc aggaccttaa agaagaaaca ggcatagatc ttgagaacat tgtttactac 1140

aaggactgca cccactattt tgtaatgaca gccaagaagc agagcctgct cgacaaaggt 1200

gtcatcatta acgactacat cgacacagag atgctgctgt gtgcggagaa cgtgaaccaa 1260

gacaacctgc tatcctatgc ccgggaagct gcagactttg ccaccaacta ccagctgcca 1320

tccttagact ttgccatgaa ccactatggg cagcctgatg tggccatgtt tgactttacc 1380

tgcatgtatg cctcagagaa cgcggccctg gtgcgggagc ggcaggcgca ccagctgctc 1440

gtggcccttg tgggtgacag cttgcttgag ccattttggc ccatgggtac aggctgtgcc 1500

```
cgtggcttcc tggcagcctt tgacacggca tggatggtga agagctggaa ccagggcacc   1560
cctcccctgg agctgctggc tgaaagggaa agtctctacc ggctgttacc tcagacaacc   1620
ccggagaaca tcaacaagaa ctttgagcag tacacgttgg acccagggac acggtaccca   1680
aacctcaact cacactgtgt caggccccat caggtgaagc atttgtatat cactaaggag   1740
ctggagcact accctctcga gagactgggc tcggtgagga gatctgtcaa cctctccagg   1800
aaggagtcag atatccggcc cagcaagctc ctgacctggt gccagcagca gacagagggc   1860
taccagcatg tcaacgtcac cgacctgacc acatcctggc gcagtgggtt ggccctgtgt   1920
gccatcatcc accgcttccg gcctgagctc atcaactttg actctttgaa tgaagatgat   1980
gctgtggaga acaaccagct cgcatttgat gtggccgagc gagagtttgg gatccctcca   2040
gtgaccacgg gcaaagagat ggcatctgcc caggagcctg acaagctcag catggtcatg   2100
tacctctcca agttctacga gctcttccgg ggcaccccac tgaggcccgt ggattcttgg   2160
cgcaaaaact atggagaaaa tgctgacctc agcttggcca aatcatccat ttctaataac   2220
tatctcaacc tcacatttcc aaggaagagg actccacggg tggatggtca aaccggagag   2280
aatgacatga acaaacggag acggaagggc ttcaccaacc tggacgagcc ttcaaacttt   2340
tccagccgta gcttgggctc caatcaagag tgtgggagca gtaaggaagg tggaaatcag   2400
aacaaagtca agtccatggc gaatcagctg ctggccaagt ttgaggagag cactcggaac   2460
ccctcactca tgaagcagga acgccgtgtc tcagggatag gtaagccggt cctgtgctct   2520
tcctccggcc ctcctgttca ctcttgctgc cccaagccgg aggaggccac acccagccca   2580
tcacctcctc tgaaaaggca gttcccctct gtggtcgtga cggggcacgt gctcagagag   2640
ctcaagcaag tgtctgctgg cagtgagtgc ctgagcagac cttggagagc cagagccaag   2700
tctgacctac agctgggtgg gacagaaaat ttcgctaccc tgccttctac ccgcccgagg   2760
gcgcaggctc tttccggggt gctgtggcgg ctgcagcaag tggaggaaaa gattctccag   2820
aagagggctc agaacttggc caacagggaa tttcacacaa agaacattaa ggagaaggcg   2880
gctcaccttg cctccatgtt tggacacggg gatttcccgc agaataaact gctctctaaa   2940
ggcctgtctc atactcatcc tccatctcct ccctctcgcc ttccgtctcc tgatccagct   3000
gcttcttcct ctccatcaac tgttgactct gcttctcctg ccagaaagga aaagaagtca   3060
ccttcagggt tccattttca tcccagccat ttgagaacag tgcatcctca gctgacggta   3120
gggaaagtgt ccagcggaat aggggctgca gctgaagtcc tggtcaatct gtacatgaat   3180
gatcacagac ctaaggccca ggccacctct ccagacctgg aatctatgcg aaagtcattt   3240
ccccttaacc tgggaggcag cgacacgtgt tacttctgta agaaacgtgt gtacgtgatg   3300
gaacggctga gcgccgaggg ccacttcttc caccgggagt gtttccgctg cagcatctgt   3360
```

gccaccacct tgcgcctggc cgcctacacc tttgactgcg atgaaggcaa attttactgc 3420
aagcctcact tcattcactg taaaaccaat agcaaacaac ggaagagacg ggcagagttg 3480
aagcaacaaa gagaggagga ggcaacatgg caagagcagg aagcccctcg gagagacact 3540
cccaccgaaa gttcttgcgc agtggccgcc attggcaccc tggaaggcag ccccccagtt 3600
catttcagcc ttccagtgct acacccactt cttggccacc ccatctgggg gaaggacagg 3660
agctggacag gccaagagct atctcccttg gctggagaag accgggaaaa agggagtact 3720
ggagccagga aggaagaaga gggagggcca gtgctggtaa aggagaagtt gggcctgaag 3780
aagttagtcc tcacccagga gcagaagacc atgttgttgg attggaatga ctccatccct 3840
gagagtgtgc acctcaaagc tggggagcga atttcccaga aaagtgctga gaatggtaga 3900
ggaggccgtg tgctaaaacc agtccgcccc ctgctgctcc ctagggcagc aggagagccc 3960
ctgccaaccc agagaggggc tcaggagaag atggggaccc ctgcggaaca agctcaaggg 4020
gagcgaaacg tgcctccacc caagtcccca ctgcggctca tagccaatgc catccgaagg 4080
tctctagagc ccctcctttc caactctgaa ggcgggaaga aggcctgggc caagcaagaa 4140
tccaaaactt tgcccacaca ggcctgcact cgctcattca gccttcggaa aaccaattcc 4200
aataaagacg gggaccagca ttcccctggg agaaaccagt cctcagcctt tagccctcct 4260
gaccctgccc tccgcaccca cagtttgccc aatcggccat ccaaggtctt tcctgcactt 4320
aggtccccac cctgcagcaa gattgaagaa gtccccacac tcctcgagaa agtgagtttg 4380
caagagaact tcccagatgc ttctaagcct ccaaagaaaa gaatctcact tttttcctcc 4440
ctcagactca aagacaaatc ttttgagagt ttcctccaag aatccagaca aagaaaggac 4500
atcagggacc tctttggcag ccccaagagg aaggtgctgc ctgaagatag tgcgcaggcc 4560
ctggagaagc tgctgcagcc tttcaaaagc acctccctgc gccaggcagc tcctcctcct 4620
cctcctcctc ctcctcctcc tcctcctcct cctcctcccc ctacagcggg aggtgcagac 4680
tccaagaact ttcccctcag agcacaggta acagaggctt cctcttctgc ctcttcaacc 4740
tcctcctcct ctgcagatga agaatttgat ccccagcttt ccttgcagtt aaaggagaag 4800
aagacactta gaagaagaaa gaagctagaa aaagcaatga agcagttggt taagcaagaa 4860
gaattgaaaa gactctataa ggctcaggcc atccagaggc agctggagga ggtggaggag 4920
cggcagaggg cttctgagat ccagggtgtg aggctggaga aggcgttgcg aggagaagca 4980
gattcaggca cacaggatga agcacagctt ttgcaggaat ggtttaagct ggttctggag 5040
aagaataaat taatgcgata tgagtcggag ctcctaatca tggcccagga actggaatta 5100
gaagatcatc aaagcagact ggagcagaaa ctgagagaga aaatgctcaa ggaggagagc 5160

cagaaagatg agaaggatct aaacgaagag caagaagtat tcaccgagct gatgcaagtg 5220

attgagcaaa gggacaaact cgtcgattcc ttagaggaac aacgcatcag agaaaaagcc 5280

gaggaccagc actttgaaag cttcgtattc tccagaggct gtcagctgag caggacttga 5340

ggaggcccgt agtccctctc cctggctgca cgttgggacc ggatcaggcc aagtgcacca 5400

cacaccctca tgggtctttc tgcaggattt atcatccctg gaacttaagt tataccttac 5460

tgcatttttt aaaattaaaa ttctcttgca cgcatggcag cttcccaagg ttcttccaga 5520

gattcaaatg aagaaaaccc aaagactctt tggcaattgg cagtcaactt cagccaggct 5580

ctcagactgg aggtgttgtt ggcagatgac cagcattgtt ttccctagaa agtgacacaa 5640

agacttgact ttcctgctac ttttatcatt ttccttccca attcattgag ttacatactt 5700

taagattttt gagaagctgc cttttcatta atatatccat atttgccttt ttttgtatgg 5760

atgaccagtt tccaaatgtc agaaagaagc agccgcagtt taaagattag gttaatattt 5820

aaattgtgtt tccagagaaa gaggagaaac cttgagatta ctgattacat aaagcaaata 5880

actcatatag caggtgttaa ttcaatccag ggtgaattta atttaccagg tgcatttata 5940

agccttaata taatatacat aagcaatgag agcttaatag aacatttgag ccttaatttt 6000

attttaag 6008

<210> 4
<211> 1633
<212> PRT
<213> Homo sapiens

<400> 4

Met Gly Glu Asn Glu Asp Glu Lys Gln Ala Gln Ala Gly Gln Val Phe
1 5 10 15

Glu Asn Phe Val Gln Ala Ser Thr Cys Lys Gly Thr Leu Gln Ala Phe
20 25 30

Asn Ile Leu Thr Arg His Leu Asp Leu Asp Pro Leu Asp His Arg Asn
35 40 45

Phe Tyr Ser Lys Leu Lys Ser Lys Val Thr Thr Trp Lys Ala Lys Ala
50 55 60

Leu Trp Tyr Lys Leu Asp Lys Arg Gly Ser His Lys Glu Tyr Lys Arg
65 70 75 80

Gly Lys Ser Cys Thr Asn Thr Lys Cys Leu Ile Val Gly Gly Gly Pro
85 90 95

Cys Gly Leu Arg Thr Ala Ile Glu Leu Ala Tyr Leu Gly Ala Lys Val
100 105 110

Val Val Val Glu Lys Arg Asp Ser Phe Ser Arg Asn Asn Val Leu His
115 120 125

Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala Lys Lys
130 135 140

Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His Ile Ser Ile Arg
145 150 155 160

Gln Leu Gln Leu Ile Leu Phe Lys Val Ala Leu Met Leu Gly Val Glu
165 170 175

Ile His Val Asn Val Glu Phe Val Lys Val Leu Glu Pro Pro Glu Asp
180 185 190

Gln Glu Asn Gln Lys Ile Gly Trp Arg Ala Glu Phe Leu Pro Thr Asp
195 200 205

His Ser Leu Ser Glu Phe Glu Phe Asp Val Ile Ile Gly Ala Asp Gly
210 215 220

Arg Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg Gly Lys
225 230 235 240

Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Ser Thr Ala
245 250 255

Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln
260 265 270

Lys Phe Phe Gln Asp Leu Lys Glu Glu Thr Gly Ile Asp Leu Glu Asn
275 280 285

Ile Val Tyr Tyr Lys Asp Cys Thr His Tyr Phe Val Met Thr Ala Lys
290 295 300

Lys Gln Ser Leu Leu Asp Lys Gly Val Ile Ile Asn Asp Tyr Ile Asp
305 310 315 320

Thr Glu Met Leu Leu Cys Ala Glu Asn Val Asn Gln Asp Asn Leu Leu
325 330 335

Ser Tyr Ala Arg Glu Ala Ala Asp Phe Ala Thr Asn Tyr Gln Leu Pro
340 345 350

Ser Leu Asp Phe Ala Met Asn His Tyr Gly Gln Pro Asp Val Ala Met
355 360 365

Phe Asp Phe Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg
370 375 380

Glu Arg Gln Ala His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu
385 390 395 400

Leu Glu Pro Phe Trp Pro Met Gly Thr Gly Cys Ala Arg Gly Phe Leu
405 410 415

Ala Ala Phe Asp Thr Ala Trp Met Val Lys Ser Trp Asn Gln Gly Thr
420 425 430

Pro Pro Leu Glu Leu Leu Ala Glu Arg Glu Ser Leu Tyr Arg Leu Leu
435 440 445

Pro Gln Thr Thr Pro Glu Asn Ile Asn Lys Asn Phe Glu Gln Tyr Thr
450 455 460

Leu Asp Pro Gly Thr Arg Tyr Pro Asn Leu Asn Ser His Cys Val Arg
465 470 475 480

Pro His Gln Val Lys His Leu Tyr Ile Thr Lys Glu Leu Glu His Tyr
485 490 495

Pro Leu Glu Arg Leu Gly Ser Val Arg Arg Ser Val Asn Leu Ser Arg
500 505 510

Lys Glu Ser Asp Ile Arg Pro Ser Lys Leu Leu Thr Trp Cys Gln Gln
515 520 525

Gln Thr Glu Gly Tyr Gln His Val Asn Val Thr Asp Leu Thr Thr Ser
530 535 540

Trp Arg Ser Gly Leu Ala Leu Cys Ala Ile Ile His Arg Phe Arg Pro
545 550 555 560

Glu Leu Ile Asn Phe Asp Ser Leu Asn Glu Asp Asp Ala Val Glu Asn
565 570 575

Asn Gln Leu Ala Phe Asp Val Ala Glu Arg Glu Phe Gly Ile Pro Pro

580 585 590

Val Thr Thr Gly Lys Glu Met Ala Ser Ala Gln Glu Pro Asp Lys Leu
595 600 605

Ser Met Val Met Tyr Leu Ser Lys Phe Tyr Glu Leu Phe Arg Gly Thr
610 615 620

Pro Leu Arg Pro Val Asp Ser Trp Arg Lys Asn Tyr Gly Glu Asn Ala
625 630 635 640

Asp Leu Ser Leu Ala Lys Ser Ser Ile Ser Asn Asn Tyr Leu Asn Leu
645 650 655

Thr Phe Pro Arg Lys Arg Thr Pro Arg Val Asp Gly Gln Thr Gly Glu
660 665 670

Asn Asp Met Asn Lys Arg Arg Arg Lys Gly Phe Thr Asn Leu Asp Glu
675 680 685

Pro Ser Asn Phe Ser Ser Arg Ser Leu Gly Ser Asn Gln Glu Cys Gly
690 695 700

Ser Ser Lys Glu Gly Gly Asn Gln Asn Lys Val Lys Ser Met Ala Asn
705 710 715 720

Gln Leu Leu Ala Lys Phe Glu Glu Ser Thr Arg Asn Pro Ser Leu Met
725 730 735

Lys Gln Glu Arg Arg Val Ser Gly Ile Gly Lys Pro Val Leu Cys Ser
740 745 750

Ser Ser Gly Pro Pro Val His Ser Cys Cys Pro Lys Pro Glu Glu Ala
755 760 765

Thr Pro Ser Pro Ser Pro Pro Leu Lys Arg Gln Phe Pro Ser Val Val
770 775 780

Val Thr Gly His Val Leu Arg Glu Leu Lys Gln Val Ser Ala Gly Ser
785 790 795 800

Glu Cys Leu Ser Arg Pro Trp Arg Ala Arg Ala Lys Ser Asp Leu Gln
805 810 815

Leu Gly Gly Thr Glu Asn Phe Ala Thr Leu Pro Ser Thr Arg Pro Arg
820 825 830

```
Ala Gln Ala Leu Ser Gly Val Leu Trp Arg Leu Gln Gln Val Glu Glu
        835                 840                 845

Lys Ile Leu Gln Lys Arg Ala Gln Asn Leu Ala Asn Arg Glu Phe His
    850                 855                 860

Thr Lys Asn Ile Lys Glu Lys Ala Ala His Leu Ala Ser Met Phe Gly
865                 870                 875                 880

His Gly Asp Phe Pro Gln Asn Lys Leu Leu Ser Lys Gly Leu Ser His
                885                 890                 895

Thr His Pro Pro Ser Pro Pro Ser Arg Leu Pro Ser Pro Asp Pro Ala
            900                 905                 910

Ala Ser Ser Ser Pro Ser Thr Val Asp Ser Ala Ser Pro Ala Arg Lys
        915                 920                 925

Glu Lys Lys Ser Pro Ser Gly Phe His Phe His Pro Ser His Leu Arg
    930                 935                 940

Thr Val His Pro Gln Leu Thr Val Gly Lys Val Ser Ser Gly Ile Gly
945                 950                 955                 960

Ala Ala Ala Glu Val Leu Val Asn Leu Tyr Met Asn Asp His Arg Pro
                965                 970                 975

Lys Ala Gln Ala Thr Ser Pro Asp Leu Glu Ser Met Arg Lys Ser Phe
            980                 985                 990

Pro Leu Asn Leu Gly Gly Ser Asp  Thr Cys Tyr Phe Cys  Lys Lys Arg
        995                 1000                 1005

Val Tyr  Val Met Glu Arg Leu  Ser Ala Glu Gly His  Phe Phe His
    1010                 1015                 1020

Arg Glu  Cys Phe Arg Cys Ser  Ile Cys Ala Thr Thr  Leu Arg Leu
    1025                 1030                 1035

Ala Ala  Tyr Thr Phe Asp Cys  Asp Glu Gly Lys Phe  Tyr Cys Lys
    1040                 1045                 1050

Pro His  Phe Ile His Cys Lys  Thr Asn Ser Lys Gln  Arg Lys Arg
    1055                 1060                 1065
```

Arg Ala Glu Leu Lys Gln Gln Arg Glu Glu Glu Ala Thr Trp Gln
1070 1075 1080

Glu Gln Glu Ala Pro Arg Arg Asp Thr Pro Thr Glu Ser Ser Cys
1085 1090 1095

Ala Val Ala Ala Ile Gly Thr Leu Glu Gly Ser Pro Pro Val His
1100 1105 1110

Phe Ser Leu Pro Val Leu His Pro Leu Leu Gly Met Leu Leu Asp
1115 1120 1125

Trp Asn Asp Ser Ile Pro Glu Ser Val His Leu Lys Ala Gly Glu
1130 1135 1140

Arg Ile Ser Gln Lys Ser Ala Glu Asn Gly Arg Gly Gly Arg Val
1145 1150 1155

Leu Lys Pro Val Arg Pro Leu Leu Leu Pro Arg Ala Ala Gly Glu
1160 1165 1170

Pro Leu Pro Thr Gln Arg Gly Ala Gln Glu Lys Met Gly Thr Pro
1175 1180 1185

Ala Glu Gln Ala Gln Gly Glu Arg Asn Val Pro Pro Pro Lys Ser
1190 1195 1200

Pro Leu Arg Leu Ile Ala Asn Ala Ile Arg Arg Ser Leu Glu Pro
1205 1210 1215

Leu Leu Ser Asn Ser Glu Gly Gly Lys Lys Ala Trp Ala Lys Gln
1220 1225 1230

Glu Ser Lys Thr Leu Pro Thr Gln Ala Cys Thr Arg Ser Phe Ser
1235 1240 1245

Leu Arg Lys Thr Asn Ser Asn Lys Asp Gly Asp Gln His Ser Pro
1250 1255 1260

Gly Arg Asn Gln Ser Ser Ala Phe Ser Pro Pro Asp Pro Ala Leu
1265 1270 1275

Arg Thr His Ser Leu Pro Asn Arg Pro Ser Lys Val Phe Pro Ala
1280 1285 1290

Leu Arg Ser Pro Pro Cys Ser Lys Ile Glu Glu Val Pro Thr Leu
1295 1300 1305

Leu Glu Lys Val Ser Leu Gln Glu Asn Phe Pro Asp Ala Ser Lys
1310 1315 1320

Pro Pro Lys Lys Arg Ile Ser Leu Phe Ser Ser Leu Arg Leu Lys
1325 1330 1335

Asp Lys Ser Phe Glu Ser Phe Leu Gln Glu Ser Arg Gln Arg Lys
1340 1345 1350

Asp Ile Arg Asp Leu Phe Gly Ser Pro Lys Arg Lys Val Leu Pro
1355 1360 1365

Glu Asp Ser Ala Gln Ala Leu Glu Lys Leu Leu Gln Pro Phe Lys
1370 1375 1380

Ser Thr Ser Leu Arg Gln Ala Ala Pro Pro Pro Pro Pro Pro Pro
1385 1390 1395

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Thr Ala Gly Gly Ala
1400 1405 1410

Asp Ser Lys Asn Phe Pro Leu Arg Ala Gln Val Thr Glu Ala Ser
1415 1420 1425

Ser Ser Ala Ser Ser Thr Ser Ser Ser Ser Ala Asp Glu Glu Phe
1430 1435 1440

Asp Pro Gln Leu Ser Leu Gln Leu Lys Glu Lys Lys Thr Leu Arg
1445 1450 1455

Arg Arg Lys Lys Leu Glu Lys Ala Met Lys Gln Leu Val Lys Gln
1460 1465 1470

Glu Glu Leu Lys Arg Leu Tyr Lys Ala Gln Ala Ile Gln Arg Gln
1475 1480 1485

Leu Glu Glu Val Glu Glu Arg Gln Arg Ala Ser Glu Ile Gln Gly
1490 1495 1500

Val Arg Leu Glu Lys Ala Leu Arg Gly Glu Ala Asp Ser Gly Thr
1505 1510 1515

Gln Asp Glu Ala Gln Leu Leu Gln Glu Trp Phe Lys Leu Val Leu

1520 1525 1530

Glu Lys Asn Lys Leu Met Arg Tyr Glu Ser Glu Leu Leu Ile Met
1535 1540 1545

Ala Gln Glu Leu Glu Leu Glu Asp His Gln Ser Arg Leu Glu Gln
1550 1555 1560

Lys Leu Arg Glu Lys Met Leu Lys Glu Glu Ser Gln Lys Asp Glu
1565 1570 1575

Lys Asp Leu Asn Glu Glu Gln Glu Val Phe Thr Glu Leu Met Gln
1580 1585 1590

Val Ile Glu Gln Arg Asp Lys Leu Val Asp Ser Leu Glu Glu Gln
1595 1600 1605

Arg Ile Arg Glu Lys Ala Glu Asp Gln His Phe Glu Ser Phe Val
1610 1615 1620

Phe Ser Arg Gly Cys Gln Leu Ser Arg Thr
1625 1630

<210> 5
<211> 8977
<212> DNA
<213> Homo sapiens

<400> 5
atggaggaga ggaagcatga gaccatgaac ccagctcatg tcctctttga ccggtttgtc 60

caggccacca cctgcaaggg aaccctcaag gctttccagg agctctgtga ccacctggaa 120

ctaaagccaa aggactaccg ctccttctat cacaagctca agtccaagct taactactgg 180

aaagccaaag ccctctgggc aaaattggac aaacggggca gtcacaaaga ctacaaaaag 240

ggaaaagcgt gcactaacac caagtgtctc atcattgggg ctggcccctg tggtctccgt 300

acagccatcg acttatcctt actgggggcc aaggtggttg ttattgagaa acgagatgcc 360

ttctcccgca acaacgtctt gcatctctgg ccattcacca tacatgatct acgaggtctg 420

ggtgccaaga agttctatgg caagttctgt gctggagcca tcgaccatat cagtatccgt 480

cagctccaac taatactttt gaaagtagcc ttgatcctag gcattgaaat ccacgtcaat 540

gtggaattcc aaggacttat acagcctcct gaggaccaag agaatgaacg gataggctgg 600

cgggcactgg tgcaccccaa gactcatcct gtgtcagagt atgaatttga agtgatcatc 660

ggtggggatg gtcggaggaa caccttggaa gggtttcgtc ggaaagaatt ccgtggcaaa 720

ctggccatcg ccatcacggc aaattttatc aaccgaaata caacagcaga agctaaagtg 780
gaagagatca gtggtgtggc ttttatattc aaccaaaaat ttttccagga actgagggaa 840
gccacaggta ttgacttgga gaacatcgtt tactacaaag atgacacaca ctatttcgtt 900
atgacagcca aaaagcagag tttgctggac aaaggagtga tactacatga ctacgccgac 960
acagagctcc tgctttcccg agaaaacgtg gaccaggagg ctctgctcag ctatgccagg 1020
gaggcggcag acttctctac ccagcagcag ctgccgtctc tggattttgc catcaatcac 1080
tatgggcagc ccgatgtggc catgtttgac ttcacttgta tgtatgcctc cgagaacgcc 1140
gccttggtgc gggagcagaa cggacaccag ttactagtgg ctctggtcgg ggacagcctc 1200
ctagagcctt tctggccaat gggaacagga atagcccggg gctttctagc tgctatggac 1260
tctgcctgga tggtccgaag ttggtctcta ggaacgagcc ctttggaagt gctggcagag 1320
agggaaagta tttacaggtt gctgcctcag accacccctg agaatgtgag taagaacttc 1380
agccagtaca gtatcgaccc tgtcactcgg tatcccaata tcaacgtcaa cttcctccgg 1440
ccaagccagg tgcgccattt atatgatact ggcgaaacaa aagatattca cctggaaatg 1500
gagagcctgg tgaattcccg aaccaccccc aaattgactc gcaatgagtc tgtagctcgt 1560
tcaagcaaac tgctgggttg gtgccagagg cagacagatg gctatgcagg ggtaaacgtg 1620
acagatctca ccatgtcctg gaaaagtggc ttggcccttt gtgcaattat ccatagatac 1680
cgccctgacc tgatagattt tgattctttg gatgagcaaa atgtggagaa gaataaccaa 1740
ctggcctttg acattgctga gaaggaattg ggcatttctc ccatcatgac aggcaaagaa 1800
atggcctccg tgggggagcc tgataagctg tccatggtga tgtacctgac tcagttctac 1860
gagatgttta aggactccct cccctctagc gacaccttgg acctaaatgc cgaggagaaa 1920
gcagtcctga tagccagcac cagatcccct atctccttcc taagcaaact tggccagacc 1980
atctctcgga agcgttctcc caaggataaa aaggaaaagg acttggatgg tgctgggaag 2040
aggagaaaga ccagtcaatc agaggaggag gaagctcctc ggggccacag aggagaaaga 2100
ccgaccctgg tgagcactct gacagacagg aggatggacg ttgccgttgg gaaccagaac 2160
aaagtgaagt acatggcgac ccagctgctg gccaaatttg aagagaatgc gcccgcacag 2220
tccatcggca tacggagaca gctgacacaa gagcgtgggg ccagccagcc gtcctgctgc 2280
ctgcctgggc aggttcgccc tgccccacc ccccggtgga aacagggctc catgaagaag 2340
gagttcccgc agaacctggg aggcagcgac acatgctact tctgccagaa gcgggtctac 2400
gtgatggaga ggctgagtgc cgagggcaag ttcttccacc ggagctgctt caagtgcgag 2460
tactgcgcca ccaccctgcg cctctcggcc tacgcctacg acatcgagga tggtaaattc 2520
tactgtaagc cacactactg ctatcgactc tctggctacg cacaaaggaa gagaccggca 2580

gtggctcccc tgtctggaaa ggaggccaaa ggacccctgc aggatggcgc caccacagat 2640
gcaaacggac gggccaacgc cgtggccagc tccactgaga gaacccccagg ttcaggcgtg 2700
aacggcctgg aggagcccag catcgccaag cgactgaggg gcaccccaga gcggatcgag 2760
ctggagaact accgcctgtc cctgaggcag gctgaggcac tgcaggaggt accggaggag 2820
actcaggccg agcacaacct gagcagcgtg ctggacacgg gcgccgagga ggacgtcgcc 2880
agcaggtcag cacgcagggc tgcagggcgc ccacccgcca cacggcccga agagtccagt 2940
gaagccggga accagaggct ccagcaggtc atgcacgcgg cggatcctct ggagatccag 3000
gctgacgtgc actggactca tatccgtgag agagaggagg aagagaggat ggcgccggcc 3060
tctgagtcct ctgcttccgg agccccattg gatgagaatg acctagagga agatgtggac 3120
tcagaaccag ccgagataga aggggaggca gcagaggatg gggacccagg ggacactggt 3180
gctgagctgg atgatgatca gcactggtct gacagcccgt cggatgctga cagagagctg 3240
cgtttgccgt gcccagctga gggggaagca gagctggagc tgagggtgtc ggaagatgag 3300
gagaagctgc ccgcctcacc gaagcaccaa gagagaggtc cctcccaagc caccagcccc 3360
atccggtctc cccaggaatc agctcttctg ttcattccag tccacagccc ctcaacagag 3420
gggccccaac tcccacctgt ccctgccgcc acccaggaga aatcacctga ggagcgcctt 3480
ttccctgagc ctttgctccc caaagagaag cccaaagctg atgcccccctc ggatctgaaa 3540
gctgtgcact ctcccatccg atcacagcca gtgaccctgc cagaagctag gactcctgtc 3600
tcaccaggga gcccgcagcc ccggccaccc gtggcggcct ccacgccccc acccagccca 3660
ctccccatct gctcccagcc ccagccttcc accgaggcca ctgtcccatc ccctacccag 3720
tcccccatac gcttccagcc tgccccggcc aaaacatcca ccccactggc ccctctccct 3780
gtccaaagcc aaagtgacac caaggacaga ctgggcagcc cccttgctgt ggatgaggcc 3840
ctcagacgga gcgacctggt ggaggagttc tggatgaaga gtgcggagat ccgccgcagc 3900
ctcgggctca cacctgtgga ccgcagcaag gggcccgagc ccagcttccc cacgcctgcc 3960
ttcaggccag tgtccctcaa atcctattcc gttgaaaagt ccccccagga tgagggactc 4020
caccttctca agcctctgtc catccccaaa aggctgggcc tgccaaagcc ggaaggcgag 4080
ccgttgtccc tgccaacccc ccggtccccg tccgacagag agctacgcag cgcccaggag 4140
gagcgcaggg agctgtccag cagctctggc ctgggcctgc acgggagctc ctccaacatg 4200
aagacactgg gcagccagag cttcaacacc tcggactccg ccatgctcac gccccccctcc 4260
agcccgcccc caccgccacc cccgggcgag gagcccgcca ccttgcggag gaagctcagg 4320
gaggccgagc ccaatgcctc ggtggtcccg ccgcccttgc ccgccacctg gatgcggccc 4380

ccccgggagc ctgctcagcc ccccagagag gaggtgcgga agtcgtttgt ggagagtgtg 4440
gaggagattc cctttgctga tgatgtggag gacacctatg acgacaagac tgaggactca 4500
agcctgcagg agaaattctt cacgcccccg tcctgctggc cgcgccccga gaagcctcgc 4560
cacccgcccc tggccaagga gaacgggagg ctgcctgctc tggaggggac gctgcagcca 4620
cagaagaggg ggctgccctt ggtgtccgcg gaagccaagg agttggccga ggagcgcatg 4680
cgagccaggg agaagtccgt gaagagccag gcgctgcggg atgccatggc caggcagctg 4740
agcaggatgc agcagatgga gctggcctca ggcgccccca ggccccgcaa ggcgtcctca 4800
gcaccctccc agggcaagga gcgccggcct gactccccca cacgccccac tctcaggggc 4860
tccgaggagc ccaccctgaa gcatgaagcc accagcgagg aggtcctctc cccgccgtcg 4920
gactcagggg gcccagatgg ctctttcact tcatccgagg gctccagtgg gaagagcaag 4980
aagaggtcgt cactcttctc cccccgcaga aacaagaagg agaagaagtc caaaggcgag 5040
ggccggcccc cggagaagcc cagctccaac ctcctggaag aagccgccgc caaacccaag 5100
tccctgtgga agtccgtctt ctccgggtac aagaaggaca agaagaagaa ggccgacgac 5160
aagtcctgcc ccagcacccc ctccagcggg gccacggtgg actctggaaa gcacagggtg 5220
cttcccgtcg taagggcaga gctgcagctc cggcgccagc tgagcttctc cgaggactca 5280
gacctctcca gcgacgatgt ccttgagaag tcctcacaga agtcccggcg agagccaaga 5340
acctacacgg aggaggaact gaatgccaag ctgacccggc gtgtgcaaaa ggcagctcgg 5400
agacaggcca agcaggagga gcttaagcgg ctgcatcgag cccagatcat ccagcggcag 5460
ctgcagcagg tggaggagag gcagcggcgg ctggaggaaa ggggcgtggc tgtggagaag 5520
gcgctccggg gcgaagcagg catgggcaag aaggacgacc ccaagctgat gcaggagtgg 5580
ttcaagctag tgcaggagaa gaacgccatg gtgcgctacg agtcggagct gatgatcttt 5640
gcccgggagc tggagctgga agaccggcag agtcgactgc agcaggagct ccgggaacgc 5700
atggcagtgg aagatcacct taagactgag gaggagctgt cagaagagaa gcagattctc 5760
aatgagatgc tggaggtggt ggagcagaga gactcactgg tggcgctgct ggaggagcag 5820
cggctccggg agagagagga ggacaaggac ctggaggctg ccatgctgtc caagggcttc 5880
agccttaact ggtcctgagc tcccacccaa cgctccattt tctgttggca tctgcctggc 5940
caggcagtgg catccaaacc acccggagcc gcgatctgag gaggcctggc acctccttgg 6000
agtttacgct cagatgcccg tgtgctgctt ggaaagtggt cgagtcccgc gtgcagtggg 6060
gagccccagg tgacagtggt tatctgagac ggctccacct cctgggagga ggcccacctg 6120
gacctcccac tcagaggagg agcacggcgt gtatggcatg acgcagggga ccaccccgcg 6180
cgctccctga ggatgtgctg gctgtgcccc ttttttccac tggcacattt ggtaagagag 6240

ggaagctgct ccccgtcaga accacagtgc gccgtgcgag gggcactgtc ttcttcatgc 6300
tccctggagc accaccaaag aaacgtaaac aataccccac gaaagcaggg tcaggggtca 6360
gggtgcgatc gagacccagg atgggggcgt ccagtcatgc ccaccccagc atcacaggag 6420
acatggaggt gcgggcaggc tcctgaatta ttatgcaaat taggaggacg caggaggggt 6480
ctgccctcca gccgaacacc acacactgga ccctaagtgg ccaaatgcct gggccgcttg 6540
ctggctgtgg cctgaggctt gtgggttgct gcattttgct tgtagttcac aaccattttg 6600
acactggaaa atgctgactt tgggggacag gatgaggccc tacattctaa gcccccagtt 6660
ggcagacagg cattgtccct gttccacatt tatgtcggga caggagatga ccttttcctc 6720
cgtgtttttc ctgtgtttgc acgttgaaat gaagctgaca acctggcaag acgctcagcc 6780
gcttcaaacc ctttttgtca attaacttat tttttaatac ttgaaaagaa gtaacttcgt 6840
ttgtgtatct ttactagagg aactgatcac ctgcgcccgg gtgcgggagc cacagcggca 6900
tctggtgcgt cctacgcgac ctggtccggg gctgcccggt gctcctcacg tgcatctatt 6960
tattagcctt tctcttcgta tcactggcct ggctggcatc agggagctgc ccagaacccc 7020
tgtgtgggtc ctcctcacag ctttctgtcc cctcctccac ccggtgcctg cctcaccctg 7080
gcgctagacc atctggacca ctcatgtgat gagggtgcat ttccgttctg ttttgggcca 7140
ggccaacagc agagctgcca ctctcaccct cccagtgaga attccggctc tgcagaactc 7200
gcccttgtct cagtttgggg gccagggcat caccttcctc cgcacatatg ttaaagaagg 7260
tttcaggatg ggccctcatc cacacaggcc aagagggtgc aaggtgggac cctggaatca 7320
tgtggctggt gagaattccc tcctcccagc ctaagattca cccagacgga aacgcgagtg 7380
ctgcagtgga tgctgggatg caggctggtg cctctcaagc agatcagcga ctcccccctt 7440
ccttccgaag gtgacgggca cctgccttgt gccggatctt catggggaca taaagggcga 7500
gccccgaatc actagctcct atagccaaac tgttcctttc tcacggttcc gcaccagcct 7560
ggctgtgtac agctcatcaa gccactgagc taatcggggt ggggtgcttg tccatcaaag 7620
cagtcagcac atagcccagc gaggagccat ccggaccaga cccgcctgcc aggggcgctc 7680
cagctgcctg ccgcctccgt gggctggggc cagctgggag cagaggcctg gcccctccaa 7740
ggcgtcccgc agtggagcag gctgagtggc tgtgctgacc ttgggctttc catgggaacc 7800
acactgtgct tcaacttgaa cattcatccc agctgcaaag gagcaaagaa cctgtgtcat 7860
ccttgtctgt ccagaagctg ccatctctct cccatgcaca tccgggagat aacgccgcta 7920
gtggccgcca cagcgctgat tctccacctg ttctcacttg aggcaaaggt ctccttttct 7980
tttctcttta cttaaaaaag gggagaggag ggtttctaga ttccatcttc aaaccccagt 8040

cgtcccataa aattggacgt gggaaaagac cttcactgcc tgcgtggcct ttcccagacc 8100
tctcctcgaa tgccacaaaa ccggtccagc cccggcagag ccgccttcgg cccttgtagc 8160
tcctgctggc cccacaacag ggaaacagtt tgcaaagtgg cttggaagga gtgtggccta 8220
gggatctgct gtaagtggcc agacgtagca ggagagccca gtgtcacctt ctggtcctgg 8280
tcagccttaa cactgtggca tcctccacag cacacatggc agcctgcagt ctggaaggtg 8340
gacccgagcc tttgcagaga ggcgccgcag agccgcaggc ctgcgcccca gccttctgct 8400
cccgactgtg aggacaccca gttctagtag agcacttttt ttaaagctcc cattttgtaa 8460
ccactagttt gcggttgact tgagtactct ggtgacttcc tgcgtcaagc gttctcaagc 8520
tgtgagaatg tgcgcagctc caggcaggtt ttctctcgga gagttaagtc ttcccttgaa 8580
ggcagggaag caggatggat acacatatat cacacacata aaacaccagg tgcgggagca 8640
gcccagactc aaggctgact aaactggagg ctgaataccg tggaggtcca catgcagctt 8700
ccctggaggg caggccggag gcgctcccgc ccctgggctt gaggatgctg caccccgtgg 8760
gcttccaggc ctgcccagat gatgccttca ggcctctgtc cctggcggcc atcctcaggc 8820
cgattttgac cagcaatgat agactcttct taaccctttc aaaataaatt tttcagtggg 8880
acagaaagga gagttaaaaa acattttttt aaaggtggta acatctgacc cacaaaggga 8940
atgggtctgt tttatgcaaa ataaaagttt ttcaaat 8977

<210> 6
<211> 1965
<212> PRT
<213> Homo sapiens

<400> 6

Met Glu Glu Arg Lys His Glu Thr Met Asn Pro Ala His Val Leu Phe
1 5 10 15

Asp Arg Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Lys Ala Phe
20 25 30

Gln Glu Leu Cys Asp His Leu Glu Leu Lys Pro Lys Asp Tyr Arg Ser
35 40 45

Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala Lys Ala
50 55 60

Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr Lys Lys
65 70 75 80

Gly Lys Ala Cys Thr Asn Thr Lys Cys Leu Ile Ile Gly Ala Gly Pro

85 90 95

Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Leu Leu Gly Ala Lys Val
100 105 110

Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val Leu His
115 120 125

Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala Lys Lys
130 135 140

Phe Tyr Gly Lys Phe Cys Ala Gly Ala Ile Asp His Ile Ser Ile Arg
145 150 155 160

Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly Ile Glu
165 170 175

Ile His Val Asn Val Glu Phe Gln Gly Leu Ile Gln Pro Pro Glu Asp
180 185 190

Gln Glu Asn Glu Arg Ile Gly Trp Arg Ala Leu Val His Pro Lys Thr
195 200 205

His Pro Val Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly Asp Gly
210 215 220

Arg Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg Gly Lys
225 230 235 240

Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Thr Thr Ala
245 250 255

Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln
260 265 270

Lys Phe Phe Gln Glu Leu Arg Glu Ala Thr Gly Ile Asp Leu Glu Asn
275 280 285

Ile Val Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys
290 295 300

Lys Gln Ser Leu Leu Asp Lys Gly Val Ile Leu His Asp Tyr Ala Asp
305 310 315 320

Thr Glu Leu Leu Leu Ser Arg Glu Asn Val Asp Gln Glu Ala Leu Leu
325 330 335

Ser Tyr Ala Arg Glu Ala Ala Asp Phe Ser Thr Gln Gln Gln Leu Pro
340 345 350

Ser Leu Asp Phe Ala Ile Asn His Tyr Gly Gln Pro Asp Val Ala Met
355 360 365

Phe Asp Phe Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg
370 375 380

Glu Gln Asn Gly His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu
385 390 395 400

Leu Glu Pro Phe Trp Pro Met Gly Thr Gly Ile Ala Arg Gly Phe Leu
405 410 415

Ala Ala Met Asp Ser Ala Trp Met Val Arg Ser Trp Ser Leu Gly Thr
420 425 430

Ser Pro Leu Glu Val Leu Ala Glu Arg Glu Ser Ile Tyr Arg Leu Leu
435 440 445

Pro Gln Thr Thr Pro Glu Asn Val Ser Lys Asn Phe Ser Gln Tyr Ser
450 455 460

Ile Asp Pro Val Thr Arg Tyr Pro Asn Ile Asn Val Asn Phe Leu Arg
465 470 475 480

Pro Ser Gln Val Arg His Leu Tyr Asp Thr Gly Glu Thr Lys Asp Ile
485 490 495

His Leu Glu Met Glu Ser Leu Val Asn Ser Arg Thr Thr Pro Lys Leu
500 505 510

Thr Arg Asn Glu Ser Val Ala Arg Ser Ser Lys Leu Leu Gly Trp Cys
515 520 525

Gln Arg Gln Thr Asp Gly Tyr Ala Gly Val Asn Val Thr Asp Leu Thr
530 535 540

Met Ser Trp Lys Ser Gly Leu Ala Leu Cys Ala Ile Ile His Arg Tyr
545 550 555 560

Arg Pro Asp Leu Ile Asp Phe Asp Ser Leu Asp Glu Gln Asn Val Glu
565 570 575

Lys Asn Asn Gln Leu Ala Phe Asp Ile Ala Glu Lys Glu Leu Gly Ile
580 585 590

Ser Pro Ile Met Thr Gly Lys Glu Met Ala Ser Val Gly Glu Pro Asp
595 600 605

Lys Leu Ser Met Val Met Tyr Leu Thr Gln Phe Tyr Glu Met Phe Lys
610 615 620

Asp Ser Leu Pro Ser Ser Asp Thr Leu Asp Leu Asn Ala Glu Glu Lys
625 630 635 640

Ala Val Leu Ile Ala Ser Thr Arg Ser Pro Ile Ser Phe Leu Ser Lys
645 650 655

Leu Gly Gln Thr Ile Ser Arg Lys Arg Ser Pro Lys Asp Lys Lys Glu
660 665 670

Lys Asp Leu Asp Gly Ala Gly Lys Arg Arg Lys Thr Ser Gln Ser Glu
675 680 685

Glu Glu Glu Ala Pro Arg Gly His Arg Gly Glu Arg Pro Thr Leu Val
690 695 700

Ser Thr Leu Thr Asp Arg Arg Met Asp Val Ala Val Gly Asn Gln Asn
705 710 715 720

Lys Val Lys Tyr Met Ala Thr Gln Leu Leu Ala Lys Phe Glu Glu Asn
725 730 735

Ala Pro Ala Gln Ser Ile Gly Ile Arg Arg Gln Leu Thr Gln Glu Arg
740 745 750

Gly Ala Ser Gln Pro Ser Cys Cys Leu Pro Gly Gln Val Arg Pro Ala
755 760 765

Pro Thr Pro Arg Trp Lys Gln Gly Ser Met Lys Lys Glu Phe Pro Gln
770 775 780

Asn Leu Gly Gly Ser Asp Thr Cys Tyr Phe Cys Gln Lys Arg Val Tyr
785 790 795 800

Val Met Glu Arg Leu Ser Ala Glu Gly Lys Phe Phe His Arg Ser Cys
805 810 815

Phe Lys Cys Glu Tyr Cys Ala Thr Thr Leu Arg Leu Ser Ala Tyr Ala
820 825 830

Tyr Asp Ile Glu Asp Gly Lys Phe Tyr Cys Lys Pro His Tyr Cys Tyr
835 840 845

Arg Leu Ser Gly Tyr Ala Gln Arg Lys Arg Pro Ala Val Ala Pro Leu
850 855 860

Ser Gly Lys Glu Ala Lys Gly Pro Leu Gln Asp Gly Ala Thr Thr Asp
865 870 875 880

Ala Asn Gly Arg Ala Asn Ala Val Ala Ser Ser Thr Glu Arg Thr Pro
885 890 895

Gly Ser Gly Val Asn Gly Leu Glu Glu Pro Ser Ile Ala Lys Arg Leu
900 905 910

Arg Gly Thr Pro Glu Arg Ile Glu Leu Glu Asn Tyr Arg Leu Ser Leu
915 920 925

Arg Gln Ala Glu Ala Leu Gln Glu Val Pro Glu Glu Thr Gln Ala Glu
930 935 940

His Asn Leu Ser Ser Val Leu Asp Thr Gly Ala Glu Glu Asp Val Ala
945 950 955 960

Ser Arg Ser Ala Arg Arg Ala Ala Gly Arg Pro Pro Ala Thr Arg Pro
965 970 975

Glu Glu Ser Ser Glu Ala Gly Asn Gln Arg Leu Gln Gln Val Met His
980 985 990

Ala Ala Asp Pro Leu Glu Ile Gln Ala Asp Val His Trp Thr His Ile
995 1000 1005

Arg Glu Arg Glu Glu Glu Glu Arg Met Ala Pro Ala Ser Glu Ser
1010 1015 1020

Ser Ala Ser Gly Ala Pro Leu Asp Glu Asn Asp Leu Glu Glu Asp
1025 1030 1035

Val Asp Ser Glu Pro Ala Glu Ile Glu Gly Glu Ala Ala Glu Asp
1040 1045 1050

Gly Asp Pro Gly Asp Thr Gly Ala Glu Leu Asp Asp Asp Gln His

1055 1060 1065

Trp Ser Asp Ser Pro Ser Asp Ala Asp Arg Glu Leu Arg Leu Pro
1070 1075 1080

Cys Pro Ala Glu Gly Glu Ala Glu Leu Glu Leu Arg Val Ser Glu
1085 1090 1095

Asp Glu Glu Lys Leu Pro Ala Ser Pro Lys His Gln Glu Arg Gly
1100 1105 1110

Pro Ser Gln Ala Thr Ser Pro Ile Arg Ser Pro Gln Glu Ser Ala
1115 1120 1125

Leu Leu Phe Ile Pro Val His Ser Pro Ser Thr Glu Gly Pro Gln
1130 1135 1140

Leu Pro Pro Val Pro Ala Ala Thr Gln Glu Lys Ser Pro Glu Glu
1145 1150 1155

Arg Leu Phe Pro Glu Pro Leu Leu Pro Lys Glu Lys Pro Lys Ala
1160 1165 1170

Asp Ala Pro Ser Asp Leu Lys Ala Val His Ser Pro Ile Arg Ser
1175 1180 1185

Gln Pro Val Thr Leu Pro Glu Ala Arg Thr Pro Val Ser Pro Gly
1190 1195 1200

Ser Pro Gln Pro Arg Pro Pro Val Ala Ala Ser Thr Pro Pro Pro
1205 1210 1215

Ser Pro Leu Pro Ile Cys Ser Gln Pro Gln Pro Ser Thr Glu Ala
1220 1225 1230

Thr Val Pro Ser Pro Thr Gln Ser Pro Ile Arg Phe Gln Pro Ala
1235 1240 1245

Pro Ala Lys Thr Ser Thr Pro Leu Ala Pro Leu Pro Val Gln Ser
1250 1255 1260

Gln Ser Asp Thr Lys Asp Arg Leu Gly Ser Pro Leu Ala Val Asp
1265 1270 1275

Glu Ala Leu Arg Arg Ser Asp Leu Val Glu Glu Phe Trp Met Lys
1280 1285 1290

Ser Ala Glu Ile Arg Arg Ser Leu Gly Leu Thr Pro Val Asp Arg
1295 1300 1305

Ser Lys Gly Pro Glu Pro Ser Phe Pro Thr Pro Ala Phe Arg Pro
1310 1315 1320

Val Ser Leu Lys Ser Tyr Ser Val Glu Lys Ser Pro Gln Asp Glu
1325 1330 1335

Gly Leu His Leu Leu Lys Pro Leu Ser Ile Pro Lys Arg Leu Gly
1340 1345 1350

Leu Pro Lys Pro Glu Gly Glu Pro Leu Ser Leu Pro Thr Pro Arg
1355 1360 1365

Ser Pro Ser Asp Arg Glu Leu Arg Ser Ala Gln Glu Glu Arg Arg
1370 1375 1380

Glu Leu Ser Ser Ser Ser Gly Leu Gly Leu His Gly Ser Ser Ser
1385 1390 1395

Asn Met Lys Thr Leu Gly Ser Gln Ser Phe Asn Thr Ser Asp Ser
1400 1405 1410

Ala Met Leu Thr Pro Pro Ser Ser Pro Pro Pro Pro Pro Pro Pro
1415 1420 1425

Gly Glu Glu Pro Ala Thr Leu Arg Arg Lys Leu Arg Glu Ala Glu
1430 1435 1440

Pro Asn Ala Ser Val Val Pro Pro Pro Leu Pro Ala Thr Trp Met
1445 1450 1455

Arg Pro Pro Arg Glu Pro Ala Gln Pro Pro Arg Glu Glu Val Arg
1460 1465 1470

Lys Ser Phe Val Glu Ser Val Glu Glu Ile Pro Phe Ala Asp Asp
1475 1480 1485

Val Glu Asp Thr Tyr Asp Asp Lys Thr Glu Asp Ser Ser Leu Gln
1490 1495 1500

Glu Lys Phe Phe Thr Pro Pro Ser Cys Trp Pro Arg Pro Glu Lys
1505 1510 1515

Pro Arg His Pro Pro Leu Ala Lys Glu Asn Gly Arg Leu Pro Ala
1520 1525 1530

Leu Glu Gly Thr Leu Gln Pro Gln Lys Arg Gly Leu Pro Leu Val
1535 1540 1545

Ser Ala Glu Ala Lys Glu Leu Ala Glu Glu Arg Met Arg Ala Arg
1550 1555 1560

Glu Lys Ser Val Lys Ser Gln Ala Leu Arg Asp Ala Met Ala Arg
1565 1570 1575

Gln Leu Ser Arg Met Gln Gln Met Glu Leu Ala Ser Gly Ala Pro
1580 1585 1590

Arg Pro Arg Lys Ala Ser Ser Ala Pro Ser Gln Gly Lys Glu Arg
1595 1600 1605

Arg Pro Asp Ser Pro Thr Arg Pro Thr Leu Arg Gly Ser Glu Glu
1610 1615 1620

Pro Thr Leu Lys His Glu Ala Thr Ser Glu Glu Val Leu Ser Pro
1625 1630 1635

Pro Ser Asp Ser Gly Gly Pro Asp Gly Ser Phe Thr Ser Ser Glu
1640 1645 1650

Gly Ser Ser Gly Lys Ser Lys Lys Arg Ser Ser Leu Phe Ser Pro
1655 1660 1665

Arg Arg Asn Lys Lys Glu Lys Lys Ser Lys Gly Glu Gly Arg Pro
1670 1675 1680

Pro Glu Lys Pro Ser Ser Asn Leu Leu Glu Glu Ala Ala Ala Lys
1685 1690 1695

Pro Lys Ser Leu Trp Lys Ser Val Phe Ser Gly Tyr Lys Lys Asp
1700 1705 1710

Lys Lys Lys Lys Ala Asp Asp Lys Ser Cys Pro Ser Thr Pro Ser
1715 1720 1725

Ser Gly Ala Thr Val Asp Ser Gly Lys His Arg Val Leu Pro Val
1730 1735 1740

Val Arg Ala Glu Leu Gln Leu Arg Arg Gln Leu Ser Phe Ser Glu
1745 1750 1755

Asp Ser Asp Leu Ser Ser Asp Asp Val Leu Glu Lys Ser Ser Gln
1760 1765 1770

Lys Ser Arg Arg Glu Pro Arg Thr Tyr Thr Glu Glu Glu Leu Asn
1775 1780 1785

Ala Lys Leu Thr Arg Arg Val Gln Lys Ala Ala Arg Arg Gln Ala
1790 1795 1800

Lys Gln Glu Glu Leu Lys Arg Leu His Arg Ala Gln Ile Ile Gln
1805 1810 1815

Arg Gln Leu Gln Gln Val Glu Glu Arg Gln Arg Arg Leu Glu Glu
1820 1825 1830

Arg Gly Val Ala Val Glu Lys Ala Leu Arg Gly Glu Ala Gly Met
1835 1840 1845

Gly Lys Lys Asp Asp Pro Lys Leu Met Gln Glu Trp Phe Lys Leu
1850 1855 1860

Val Gln Glu Lys Asn Ala Met Val Arg Tyr Glu Ser Glu Leu Met
1865 1870 1875

Ile Phe Ala Arg Glu Leu Glu Leu Glu Asp Arg Gln Ser Arg Leu
1880 1885 1890

Gln Gln Glu Leu Arg Glu Arg Met Ala Val Glu Asp His Leu Lys
1895 1900 1905

Thr Glu Glu Glu Leu Ser Glu Glu Lys Gln Ile Leu Asn Glu Met
1910 1915 1920

Leu Glu Val Val Glu Gln Arg Asp Ser Leu Val Ala Leu Leu Glu
1925 1930 1935

Glu Gln Arg Leu Arg Glu Arg Glu Glu Asp Lys Asp Leu Glu Ala
1940 1945 1950

Ala Met Leu Ser Lys Gly Phe Ser Leu Asn Trp Ser
1955 1960 1965

<210> 7

<211> 14172
<212> DNA
<213> Drosophila

<400> 7
atgagccgcc aacaccagcg gcaccaccag cagcatcacc acctgccgcc gcaccagcaa 60

ccgcagcagc agatgccgca acaacagcag cagctgacgg cgcagcagca gcaacaacag 120

cagctgctga tggcggagca cgcggcggcc gcggaggcgg cggagctatt cgacctgctg 180

tgcgtggcca caacgatgcg ccagatcctg gcgctccatc gggccatgtg cgaggctgtg 240

ggattgagac cctcgcctct gaacgacttc tacccacggc taaaggccaa ggtgcgttcg 300

tggaaggcgc aggccctgtg gaagaagttc gacgccagag ctgcccatag agtctacggc 360

aagggagctg cctgtactgg cacacgcgtc ctggtcatcg gagcagggcc ctgtggactg 420

cgcaccgcca tcgaggccca actgctgggc gccaaggtgg tggtgctgga gaaacgcgat 480

cgcatcaccc ggaacaatgt gctccatctg tggccattcg tcatcacgga tctgcgcaac 540

ttgggcgcaa agaagttcta cggcaagttt tgcgccggct ccatcgatca catctccatt 600

cggcagctgc agtgcatgct gctcaaggtg gcgctgctcc tgggcgtaga gatccacgag 660

ggagtcagtt ttgatcacgc tgtagagccc tctggcgatg gcggcggatg gagggcagct 720

gttactcccg cagatcatcc tgtatctcac tacgaattcg atgtgttgat cggagcggat 780

ggcaagcgga atatgctgga ctttaggagg aaggagttcc gcgggaagct ggccatcgct 840

attacagcga actttatcaa caagaagacg gaggcggagg ctaaagtaga ggagatcagt 900

ggggtggctt tcatcttcaa ccaggccttc ttcaaggagc tgtacgggaa gacgggcatc 960

gacctggaaa acatcgtcta ctacaaggac gagacgcact acttcgtgat gacggccaag 1020

aagcacagtc taattgacaa gggcgttatt atcgaggata tggccgatcc cggcgagctt 1080

ctcgccccag ccaatgtgga tacacaaaag ctgcacgact atgcacgcga ggctgcggag 1140

ttctccaccc aataccaaat gccaaacctg gagttcgctg ttaatcacta cggcaaacca 1200

gatgtggcca tgttcgactt cacatcgatg tttgccgccg agatgtcctg tcgggtgatt 1260

gtgcgcaaag gagctcgcct gatgcagtgc ctcgtgggtg acagtctgct cgagccgttt 1320

tggcccactg gatcgggttg tgcccgtgga ttcttatcca gcatggatgc tgcctatgcc 1380

atcaagcttt ggtccaaccc gcagaacagc acacttggcg ttctggcgca gcgcgaaagc 1440

atctaccggc tgcttaacca gaccacgccg gacaccctgc agcgggacat cagtgcctat 1500

accgtggatc cggccacgcg ctatccgaat ctgaacaggg agtcggtcaa tagctggcag 1560

gtcaaacatc tggtcgacac ggacgacccg tccattctgg agcagacctt catggacacg 1620

catgctctgc agaccccgca tttggacaca ccgggcagac gcaagcgacg cagtggagac 1680

ttgctgcccc agggtgccac gttgctgaga tggataagtg cccagctgca ttcctatcag 1740
tttattcccg aactcaagga ggcttcggat gtgttccgga atggacgcgt tctgtgtgcg 1800
cttatcaatc gctatcgtcc tgatctcatc gactacgctg ccaccaagga catgagtccc 1860
gtggagtgca atgagctgtc attcgccgtc ctagagcgcg aactccacat cgatcgcgtc 1920
atgagtgcca aacagtcgct ggacttgacc gagctggagt cgcgaatctg gctcaactat 1980
ttggaccaga tctgcgactt gtttcgcggc gagatccccc atatcaagca ccccaagatg 2040
gactttagcg atttgcgcca gaagtatcgt atcaaccata cgcatgccca acccgacttc 2100
tccaagctgc tggcaacgaa acccaaggcc aagtcgccga tgcaggatgc tgtggacata 2160
cccacgacag tgcagcggcg ctcggtgctc gaggaggagc gagccaagcg gcagcgtcgc 2220
cacgagcagc ttcttaacat cggtggaggg gcagcaggag ccgccgccgg agttgccgga 2280
agcgggacag gaaccacaac gcagggtcaa aacgatacgc cacgccggtc caagaagcgc 2340
cgtcaggttg acaaaaccgc caatattgag gagcgccagc agcgcttgca ggagatcgag 2400
gagaatcggc aggagcggat gagcaagcgg cgccagcagc gctgtcacca gacgcagaat 2460
ttctacaaga gccttcagct cctgcaggcg ggcaagctct tgagggaggg tggtgaggcg 2520
ggagtggccg aggatggcac cccattcgag gactactcga tattcctcta ccgccagcag 2580
gcgcccgtat tcaatgatcg cgtcaaggac ctcgagcgaa agcttctgtt tcccgatcgc 2640
gaacggggag atattccgtc ggcattgccg cgcacggcgg acgagcagtt cagcgatcgg 2700
attaaaaaca tggagcagcg gatgacggga cgtggtggcc ttggtggtga caagaagccc 2760
aaggatctga tgcgggctat cggcaagatc gactcgaacg attggaatgt acgcgagatc 2820
gagaagaaga tcgagctgtc gaagaagacg gagatccacg ggcctaaggg ccgcgagaag 2880
gttcccaagt ggagcaagga gcagttccag gcgcgacagc acaagatgtc caagccgcaa 2940
cgccaggatt cgcgtgaggc ggaaaagttc aaggacatcg accagactat ccgcaacttg 3000
gacaagcagc taaaggaggg ccacaatctg gatgtgggcg agcggggacg caacaaagtg 3060
gcctccattg ccggtcagtt tggcaaaaag gatgaggcca attcggatga gaagaacgcg 3120
ggcagcagca atgccaccac caacaccaac aacacagtca tacccaaatc tagttccaag 3180
gtggcgctgg cctttaaaaa gcaggctgcc tccgaaaagt gccgcttctg taagcaaacc 3240
gtctacccga tggagaagac caccgtggag ggattggttc tgcatcgcaa ttgccttaag 3300
tgccaccact gccacaccaa cttgcgtctg ggaggctacg cctttgatcg ggacgatccg 3360
cagggccgac tttactgcac ccaacacttc cggttgccac ccaaaccgct gccgcagcgc 3420
accaacaaag ccaggaaatc cgctgccgct caacccgcct cgcctgctgt accaccaact 3480
gcgggatccg tacccactgc agctgccaca tcggagcata tggacaccac tccacccagg 3540

gaccaggtgg atctactgca gacctcgcga gcaaatgcct ctgccgatgc catgtccgat 3600
gatgaagcca atgttatcga tgagcacgaa tggtctggtc gcaacttctt gcccgagtcc 3660
aacaacgatt cccaatcaga gctatccagt tcagatgaat cggatacgga atcggattcg 3720
gagatgtttg aggaggcgga tgattcgccg tttggtgctc agaccctcca gctggcgtcg 3780
gattggattg gaaagcaata ctgtgaggac agtgatgatt ctgacgattt ctacgactca 3840
agtgaagatg acggcaaaga tgacaccgag ggtgaggaat tcaagaaggc ccgcgaattg 3900
aggcgccagg aagttcgcct gcagccgttg cccgccaatc tgcccacaga tacggagacc 3960
gagaaactca agttaaacgt agacaataaa gaaaacatgg cggacggaag ctctctgaaa 4020
tcgggcaatt cctttgagtc cgcgcgcagc cagccgtcta cgcccctgtc cacgcccact 4080
cgtgtcgaga tggagcaact ggagcgagat gctccacgca agtttagcag cgaaatcgag 4140
gcaatcagcg agaaacttta ccacatgaac aacatggtaa agatgaacaa ggacctcgag 4200
gtgctggcca aggaaaatct ggtcaagagc ggcatcctgc gcaagctcac gctgaaggag 4260
aagtggctgg cggagaatgc ggccatagca gcaggccaaa aagtgactcc gactcctagt 4320
gcgactgctc ctgggcttca acccaagtcg aagttcgatg aaaagttcga aaaggtagtg 4380
agtccacctc agccggttgt cgaacccaaa cccaagcccg taatcgattt caatttggat 4440
gagttaaaac cacgtaaacc gaactttgag gagcgaccca aggagcagct gcccaggcct 4500
gaaagtttga aaaaacctcc acaacaaaaa ccgaagggca gcagcacaaa cgtgagccgc 4560
tccaacagct tgaagagtaa tgccagcaac gggagtccaa aggtgaaaaa agctcctgtt 4620
tcaaataaca gcaaaatgca aattgaagga atccttgata ccttgagaaa gattcagagc 4680
cagaacagta gcgatcagga tgaagatatg gatgtagatg aggatgtgga aagaaagcca 4740
aataaagagc tcaacagcaa gttgaaggaa atccaagcca gcagctttgc cggcacaatg 4800
gaccatatta aatcccagct gactatgcca acggttagtg cgcaggcacc accctccatg 4860
gatctttcca agtactttcc caaccagaag caagagaaga gctccaccag cagcacaaac 4920
aagaatcagg ttaccctgaa ggatgtaaat ctggccaagt acttccccag cagtccagct 4980
ccacaacgga gaaccgtgga aacggtggcc gatcgactaa agaagtcaca gactgaggct 5040
gctctagcta aaaccaagct actggaggat caggccaata accaggccga aaagaccaag 5100
aaagaggtcg aaaaggaagg agagtctaaa aaaatcacga aaaaggtggc agactccaaa 5160
gcagtacctc caaaacggca ggcctcgtta gacaccttta gcttaaggga gcaccaaatg 5220
gatggagcat tagacctgac caaaaagaag ggtcccacca aagcgagcgc tggtgttaaa 5280
aaaccagcca aatcagggag taccacatcc gtgacaaaag ccacggccac atccgaagga 5340

aagaccatta agattgtaaa gaaaatcgta cccaaaggaa ccaaagctaa aaaggcagca 5400
gaagcggccc aagaatccgc agtagtagag gctccgccag aaaagaaacc tccaaaagat 5460
gaagcagagc ggatattgga cgaaattctg ggagatggtg agtatcgttc acccagctcc 5520
gagtatcagc gcttgttcca ggacgaaaag tcgcctagcg atttgtccga taacatagat 5580
cggatactcg aggaatccga gttggatgtg gaattgggtc tgccaaaacg cagcagtaag 5640
aaactggtaa aaaccaaatc cctgggcgag ggtgactttg acatgaagcc ttcaaaagaa 5700
agacttactg gagtacagaa catacttaag cgatttgagt caatgagttc agtcacctca 5760
cagaacagtg atgaacaggc ggggttcaaa ttacgtcgca tggagtccac caccagtaat 5820
ctgagcagct tgacccgttc cagggaatca cttgtctccg tcagcgattc aatgagtgat 5880
ttggagaaaa ccatggacta cctccgtaat gaatggcgca atgaggccac caattttctg 5940
caaaagaagc gtgataaatt ctatgccaaa aaggaggagc aagaaaagga ggctaaaatt 6000
ctggctaaac ctgatccctt agacaatctt cctgttcaat atcgcgactc caagttggcc 6060
aagttctttg gcttggccgc tagcaagtca ccagaaaacc gaaagtcccc tattaagaag 6120
aagaaatctc cctctaaaac gccaaaggtg accaaagcaa acaattcact ggaggagctg 6180
gccaagatta gtaatgttcg ccagaccaag aaggcacagc cgaagactct aaagcctgta 6240
gaagtaaaac ctttaaagcc agccagtcca gttccagatg attttgagat tcttgatctg 6300
ctggaaaaag ccacggaggc caaagaactg gaacgctcga agaccaagag tccagccgta 6360
gaatctatca gtcagacgcc caaagaagct atagtagaaa tttccttacc agtggaagac 6420
attaaaaatc tacccaaaac cgggtgtgat aagtcctcaa acagctctcg ccgtggttcc 6480
caatctagtt tgatcatgtc gcgacgacac tcggagattt ctctaaacga gaaactcaac 6540
caggacgcac tcgcagcctt gaaccaaata gaaaaagaaa gagaggcgga gcaagtggac 6600
gaacttttcc agagcatggt ggaggaaatg gagcaggaac cgcagcccac agctattgtc 6660
gaacctccag aagaagacat cgatgcagat tccttgtgca caacaattag caagagtccc 6720
agtgcacagc ccgttacggt ggtcaaacgt ggcagttccg aggatcagag catagagaag 6780
cttttcagcc acttttccga cgagatgctg gtgaatgtgg agtttgactc taacgatgag 6840
cttgtaggaa ttacaccaag ggcaacactc gtttcccgaa acacagaaga tcgggactat 6900
ctggataaac tagagtcttt ggagcgcgat gaggagactt tccaaccggt tgttggggaa 6960
aaattcatac aggaaaatgt ccaggacgaa gtggacggat tgcatttccc atcacgccca 7020
caacggagac ccaaaagcag ttcatcttcc agtgaaccat cacttcctgt ggctcctcaa 7080
aggttggaaa aaaagctatc gaaattagat ccggaggata tgcctccttc tgtacaggat 7140
ctattacaac aggtttacca gaagaacatc caacctgaat tagtggaagt aattccagtt 7200

gaaggcaaac aaactttaag gttccccagc atgttggcgg aagaagatgt agacgaagtg 7260
gatcactcta aagaaggaat caagaaaatt gaaacggctc cagaagaggt tcgtaaggta 7320
accgagccgg aggatgttgc tcgggtaatt ccaagtccaa tcaaaccgtc aataagtcag 7380
agcaactccc tcaagagcga aaattcctct ggcagcagtt tagtggaaat accaaaaata 7440
attgcaccac ccaagtcgag tagcaaggaa aactcttcgg attgggacag ggagaagctg 7500
cccgcatcac ccatgccacg acgaagactt ctgccaaatc aaacgccata taaagcccca 7560
agcgtggcta gtaaggagag ctccttagag tgggacatgg agaagctacc caacagtccc 7620
atgttgccaa ggaggaacaa gatgcgagca atctcaccta gtaccaaccc tgtgcaactt 7680
ctcaacaatt taccctccga tgtagatgac gaggcggcac agagacgcct tatcgaggat 7740
ttcgagcagg agagacgtca ggctctgatc aagcgggatg agaacttcga ggccattgcg 7800
gcggagcaac gcaggcgcga ctccttacag agcagcagta actcgagcag caaacgcagt 7860
ttaccaccgc ccacgccgcc catgatggca tcgcgacgcg gtacaactca agacacaaat 7920
cgcacccagg ataccgcatc gcggcacgag ggaacgccgc ccatgttcaa gaagctggat 7980
gtcgatggca gcggtacatc aatggattcc acctcctgct ccacgcggcg cagttcgttt 8040
gcatttatag agttacagga taacaaacca gtgattgtgc ccatgcccaa gaaactgaag 8100
ttgccaaagc cggagccacc taggttcgta cccgagccag tggccactga tgagcctgtg 8160
cccgaggttt ttcagggtcg cgcttggccc aagacacagt tggaaggaga ggtcgatcta 8220
ggcgattcgg ataatgaaga tgaaacagaa aagctaaaga aacagctgcc cgaatacgct 8280
cgttcggact ctcctccttc ggctgcattc aagaatcgca agtggccgga tgggaaaaca 8340
gtttttgata aacgagccga atctcttgag gaggaagata tcttcgaagg attaccgtcc 8400
ccaaggaaaa gaggttctca aagattcatg gacaagccgc gctctcaatc gccacagcct 8460
ttcaaaccgc tggccaacag ctccaggaaa agctcaaagt cttttagcga ccttaagaaa 8520
ggaccctcct tgcaatctct gtcggcgcaa tccagccagg acacggacac actgtccacc 8580
accacaacag tagccacagc tcgtcctgct agctatgcca actacgagga ccccatggat 8640
gccagtaccc aagctttgct ggatcgaagc aaacggttac acaaccgcaa aagagatttc 8700
gtaaacgagc gagtagtgga gcgcaacccc tatatgaggg atgtgcttag gagcacggat 8760
cgccgtgatt acgacgacgt ggatgaagat ctgactagct acaggccaag acattatgcc 8820
agctccacgc taaatcgttt ccccaacacc acaataagaa aaagcaacaa ctacgattac 8880
ctcagtccca gcagtgatta tttgagcagg agaagctaca taccgagcgc cagtgcaacc 8940
agcagttact atccatcgac cacgcgtagc tcccatttga gtgacctgtt ccgacgacgc 9000

agccccgcca gcggaactgt atccgcactt tccggctacg gcaacaaaga gtcgtgcgtt 9060
atctcaatcg ggctggcctt agatcgagtt ggccacttga ttgaaagtaa atgcacctgg 9120
gtacgatcta cgaaggttca aaccgagtcc gagagcactt cacccgacga agtggagctc 9180
aattctgcca ctgagatatc caccgactct gagtttgaca acgatgagat tatacgccag 9240
gcgcccaaaa tcttcatcga tgacacccat ctaaggaagc ccaccaaggt tcagatcaag 9300
tccaccatga tcggacccaa tgcagcttcc gccggactcc atcagaagca gttggcggcg 9360
cgtgagaagg gcggcagcta cctccagaag taccaaccac aaccgccact gccacagttt 9420
agaccgttgg tccaggtgga tcccaccctg ctcattggca gccagcgcgc tcctcttcag 9480
aatccacggc caggagacta cttgctaaac aagacggcca gtacggaggg tatcgcctca 9540
aaaaagagcc tggggctaaa aaagcgctat ctgctgggtg agccggccaa tggcaataag 9600
atccagaagt ccggatccac ttcagtgctg gattcacgca ttcgcagctt ccagtcgaac 9660
atatcggagt gccagaagct tttgaatccc agcagcgaca taagtgccgg catgcgaacc 9720
ttcctcgatc gcacaaagtt gggcgaaggc agccagacga caccoggaca gacgaacgaa 9780
ctaatccgtt ccgccaccag caatgtgatt aacgatctgc gcgtggagct tcggatacag 9840
aaaactggct ccagccactc cacggacaac gagaaggaaa acgttttcgt gaactgtaag 9900
aacgagctga acaaggggat ggaatacacg gatgcggtca atgccacgct gctggaccag 9960
ctggccagaa aaagttcacc caccacgccg acgaataaga cggtggtcga ggttattgac 10020
ctggttacac ctgagaagcc aattgacatt atcgatctaa cggcactgga aacgccgaaa 10080
aagcagttgg tcgatggtag cgccatggat gtagatgaac gcctcacacc cgatagcaac 10140
aaaatcagcg aactgcagca ggaagtgaag gaggaaccca agccggatgt ctctagggat 10200
gtgaaagaat gcataccaga tatactggga cacattaagg agggaacggg atcgaaggag 10260
ccaggtggag aggaccaaca gagcctgctg gagcagtcgg acgaagagaa gcgcgactca 10320
ccggaaaagg atgtggccga acatgagctt tatgaaccgg acagtgtgca gatccaggtg 10380
cccaatatcc catgggaaaa aagcaagccg gaggtcatgt ctaccaccgg cagcagtggc 10440
tccatctgct caagctcaga ctcttctagt attgaagaca tccagcacta cattttggag 10500
tccacaacta gtccagatac tcagacagtt ggcggaaagc acaatgtgcc ccgtttggag 10560
gtgcacgaca caagtggtgc cctgatgcag gtggacagcc tgatgattgt gaacggaaag 10620
tatattgggg atcccgagga tgtcaagttc ttggatatgc cggccaatgt tattgttccg 10680
ccagcaccgg cgcttaaaac gaatgagctg gatatggagg atgaccaaga ggcggaggcg 10740
gaaccagtaa ctgctactcc ggagccggtg gaatgtacgg tcatcgaggc tgagcgccgt 10800
gttactgctc cccctccttt gccagagatg ggtccaccca aactgaagtt cgatagcaaa 10860

aatgagaaca agatcgagag cttgaagaat cttccgttga tcgtagagag caatgtggag 10920
cacagtcagg cagtgaaacc cattactctt aacttaagca atctggccag gacgccggat 10980
acaccaacca cgcccacggc gcacgatagc gataaaacac ccactgggga aattctctcg 11040
cgaggatctg actcagaaac cgagcacact ggcactggtc aggtactaac ggagacggaa 11100
ctctccgact ggacggccga cgactgtatc tcggagaact ttgttgactt ggagttcgcg 11160
cttaactcta acaagggtac gataaaacgg cgcaaggatc gacgacgcag tggagcaagc 11220
aaacttccca gtggcaacga ggtaatccac gagctggcca ggcaggcgcc agtggtgcaa 11280
atggatggaa ttcttagtgc catcgacatt gatgacatag agttcatgga cacgggttcg 11340
gagggttctt gtgctgaagc ttatcccgca acaaatacag ctctcattca gaatagaggt 11400
tacatggagt acatcgaggc ggagccgaaa aagacgaccc gcaaggcagc tccaccatcc 11460
agttacccag gaaatttacc gcctttaatg acgaagcggg acgagaaact gggcgttgat 11520
tacattgagc agggggcgta cataatgcac gatgatgcaa agacgcctgt gaatgaggtg 11580
gctcctgcca tgacccagtc gctaactgac tcaatcacgc tcaatgaact ggatgatgac 11640
agtatgataa tatcccaaac ccagccaacg acaacggagg aaagtgaggc actgacggtg 11700
gtcaccagtc cacttgacac gtcctcgccc agggttctcg atcaatttgc atccatgttg 11760
gcggcgggaa aaggtgactc cacacccagt agctcagagc aacaaccaaa gacgtctacg 11820
gtgacgagca gcagcactgg gcccaactcc tcgacaacag gaaacgtctc gaaggagccg 11880
caggaggagg acctgcaaat ccagtttgag tatgttcgag cactgcagca gcggatatcg 11940
cagatcagca cccaacggcg taagagctct aagggagagg cacctaacct gcagctaaac 12000
agtagcgcac ctgtgataga atcagccgag gatccggcca agcccgcaga ggagcctctg 12060
gtctcaatgc gaccgcggac caccagcatt tccggaaagg taccggagat acccacactt 12120
agcagcaagc tggaagagat aaccaaagaa cgcactaagc aaaaggatct gattcacgac 12180
ctagtcatgg acaagttgca gtcgaagaag cagctaaacg ctgagaagcg tctgcaccgg 12240
agtcgacagc gcagtttgct gaccagtggc tatgccagtg gctccagcct tagtccgacg 12300
cccaagctgg ctgctgcttg cagtccgcag gattccaact gctctagcca agcgcactac 12360
cacgcctcca cggcggagga ggccccgaag ccgccggcgg aaaggccgtt gcagaagtcc 12420
gccacgtcca cctatgtgtc gccttatcgc actgtccaag cgcccacacg tagtgctgat 12480
ctctataagc cgcgcccctt cagcgaacac atcgattcga acgctctggc gggttacaag 12540
ctcggcaaga cggcctcgtt taatggcggc aagttgggcg actttgcgaa acccattgcc 12600
ccggcgagag ttaaccgagg aggaggtgtc gcgaccgcgg atatagccaa tatttccgcg 12660

tcgacggaga acctaagaag cgaggccagg gccagggctc gtcttaagtc taacacagag 12720
ctgggcctta gtcccgagga aaagatgcag ctaatacgtt caagattgca ctacgaccaa 12780
aacagatctc tgaagccgaa gcaactggag gagatgccat ccggggatct ggcggcacgt 12840
gcccgcaaaa tgagtgcctc gaagagcgtc aatgatctgg cctacatggt gggacagcag 12900
cagcagcagc aggttgagaa ggatgccgtg ctccaagcca aggcggctga ctttacatcc 12960
gatcccaatt tggcgtccgg tggtcaggag aaggcaggca aaactaagtc cggacgcagg 13020
ccaaaggatc cggagcggcg taagagtctc atacagtcgt tgtccagctt cttccaaaag 13080
ggatctggat ccgcggcctc cagttccaag gagcagggcg gcgctgtggc tgccgtccac 13140
tctgaacagt cagagcgacc aggcaccagc agcagcggca cgcccacaat atcggatgcg 13200
gcgggtggag gcggaggagg aggtggcgtc ttcagcagat tccgcatctc gcccaagtcc 13260
aaggagaagt caaagtcttg ctttgatctg aggaatttcg gttttggtga caaggatatg 13320
ctggtctgca atgcagcatc tccagcagga gccacatccg catcacagaa aaatcactcg 13380
caagagtatc tgaacaccac gaacaacagt cgctatcgaa agcaaacgaa cactgcgaaa 13440
ccgaaacccg aatcgttctc ttcatccagt ccgcagctct atatacacaa gccccaccac 13500
ctggccgcag ctcatcccag tgccctggac gaccagacac caccacccat accgcctctt 13560
ccactgaatt atcagagatc cgatgatgag agctacgcta acgagacacg agagcataag 13620
aagcaacgtg ccatatcgaa ggcttcacga caagctgagc tcaagcgatt gcgaatcgct 13680
caagagattc agcgggaaca ggaggagatc gaggtgcaac tgaaggatct ggaggcacgc 13740
ggcgtgctta ttgagaaggc cttgcgaggc gaggcgcaga atattgaaaa cctggatgcg 13800
acaaaggaca acgacgagaa gctacttaag gaacttttgg agatttggcg caacatcaca 13860
gcactcaaga aacgcgatga ggaactgact ataaggcaac aggaactgca actggagtat 13920
cggcatgccc agctgaagga agagctcaat ctgcgcttgt cctgcaacaa actggacaaa 13980
agctctgccg atgtggccgc cgagggagca attctcaacg agatgctgga aattgtcgcc 14040
aagcgagccg ccctacgacc cacagcctcc cagctcgacc tcacggcagc gggatcagca 14100
tccacgtccg ccgaggcaac gggcattaag ctgacgggac aaccgcatga ccacgaagaa 14160
tcgatcattt ga 14172

<210> 8
<211> 4723
<212> PRT
<213> Drosophila

<400> 8

Met Ser Arg Gln His Gln Arg His His Gln Gln His His His Leu Pro

1 5 10 15

Pro His Gln Gln Pro Gln Gln Gln Met Pro Gln Gln Gln Gln Gln Leu
20 25 30

Thr Ala Gln Gln Gln Gln Gln Gln Gln Leu Leu Met Ala Glu His Ala
35 40 45

Ala Ala Ala Glu Ala Ala Glu Leu Phe Asp Leu Leu Cys Val Ala Thr
50 55 60

Thr Met Arg Gln Ile Leu Ala Leu His Arg Ala Met Cys Glu Ala Val
65 70 75 80

Gly Leu Arg Pro Ser Pro Leu Asn Asp Phe Tyr Pro Arg Leu Lys Ala
85 90 95

Lys Val Arg Ser Trp Lys Ala Gln Ala Leu Trp Lys Lys Phe Asp Ala
100 105 110

Arg Ala Ala His Arg Val Tyr Gly Lys Gly Ala Ala Cys Thr Gly Thr
115 120 125

Arg Val Leu Val Ile Gly Ala Gly Pro Cys Gly Leu Arg Thr Ala Ile
130 135 140

Glu Ala Gln Leu Leu Gly Ala Lys Val Val Val Leu Glu Lys Arg Asp
145 150 155 160

Arg Ile Thr Arg Asn Asn Val Leu His Leu Trp Pro Phe Val Ile Thr
165 170 175

Asp Leu Arg Asn Leu Gly Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala
180 185 190

Gly Ser Ile Asp His Ile Ser Ile Arg Gln Leu Gln Cys Met Leu Leu
195 200 205

Lys Val Ala Leu Leu Leu Gly Val Glu Ile His Glu Gly Val Ser Phe
210 215 220

Asp His Ala Val Glu Pro Ser Gly Asp Gly Gly Gly Trp Arg Ala Ala
225 230 235 240

Val Thr Pro Ala Asp His Pro Val Ser His Tyr Glu Phe Asp Val Leu
245 250 255

Ile Gly Ala Asp Gly Lys Arg Asn Met Leu Asp Phe Arg Arg Lys Glu
260 265 270

Phe Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Lys
275 280 285

Lys Thr Glu Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe
290 295 300

Ile Phe Asn Gln Ala Phe Phe Lys Glu Leu Tyr Gly Lys Thr Gly Ile
305 310 315 320

Asp Leu Glu Asn Ile Val Tyr Tyr Lys Asp Glu Thr His Tyr Phe Val
325 330 335

Met Thr Ala Lys Lys His Ser Leu Ile Asp Lys Gly Val Ile Ile Glu
340 345 350

Asp Met Ala Asp Pro Gly Glu Leu Leu Ala Pro Ala Asn Val Asp Thr
355 360 365

Gln Lys Leu His Asp Tyr Ala Arg Glu Ala Ala Glu Phe Ser Thr Gln
370 375 380

Tyr Gln Met Pro Asn Leu Glu Phe Ala Val Asn His Tyr Gly Lys Pro
385 390 395 400

Asp Val Ala Met Phe Asp Phe Thr Ser Met Phe Ala Ala Glu Met Ser
405 410 415

Cys Arg Val Ile Val Arg Lys Gly Ala Arg Leu Met Gln Cys Leu Val
420 425 430

Gly Asp Ser Leu Leu Glu Pro Phe Trp Pro Thr Gly Ser Gly Cys Ala
435 440 445

Arg Gly Phe Leu Ser Ser Met Asp Ala Ala Tyr Ala Ile Lys Leu Trp
450 455 460

Ser Asn Pro Gln Asn Ser Thr Leu Gly Val Leu Ala Gln Arg Glu Ser
465 470 475 480

Ile Tyr Arg Leu Leu Asn Gln Thr Thr Pro Asp Thr Leu Gln Arg Asp
485 490 495

Ile Ser Ala Tyr Thr Val Asp Pro Ala Thr Arg Tyr Pro Asn Leu Asn
500 505 510

Arg Glu Ser Val Asn Ser Trp Gln Val Lys His Leu Val Asp Thr Asp
515 520 525

Asp Pro Ser Ile Leu Glu Gln Thr Phe Met Asp Thr His Ala Leu Gln
530 535 540

Thr Pro His Leu Asp Thr Pro Gly Arg Arg Lys Arg Arg Ser Gly Asp
545 550 555 560

Leu Leu Pro Gln Gly Ala Thr Leu Leu Arg Trp Ile Ser Ala Gln Leu
565 570 575

His Ser Tyr Gln Phe Ile Pro Glu Leu Lys Glu Ala Ser Asp Val Phe
580 585 590

Arg Asn Gly Arg Val Leu Cys Ala Leu Ile Asn Arg Tyr Arg Pro Asp
595 600 605

Leu Ile Asp Tyr Ala Ala Thr Lys Asp Met Ser Pro Val Glu Cys Asn
610 615 620

Glu Leu Ser Phe Ala Val Leu Glu Arg Glu Leu His Ile Asp Arg Val
625 630 635 640

Met Ser Ala Lys Gln Ser Leu Asp Leu Thr Glu Leu Glu Ser Arg Ile
645 650 655

Trp Leu Asn Tyr Leu Asp Gln Ile Cys Asp Leu Phe Arg Gly Glu Ile
660 665 670

Pro His Ile Lys His Pro Lys Met Asp Phe Ser Asp Leu Arg Gln Lys
675 680 685

Tyr Arg Ile Asn His Thr His Ala Gln Pro Asp Phe Ser Lys Leu Leu
690 695 700

Ala Thr Lys Pro Lys Ala Lys Ser Pro Met Gln Asp Ala Val Asp Ile
705 710 715 720

Pro Thr Thr Val Gln Arg Arg Ser Val Leu Glu Glu Glu Arg Ala Lys
725 730 735

Arg Gln Arg Arg His Glu Gln Leu Leu Asn Ile Gly Gly Gly Ala Ala
740 745 750

Gly Ala Ala Ala Gly Val Ala Gly Ser Gly Thr Gly Thr Thr Thr Gln
755 760 765

Gly Gln Asn Asp Thr Pro Arg Arg Ser Lys Lys Arg Arg Gln Val Asp
770 775 780

Lys Thr Ala Asn Ile Glu Glu Arg Gln Gln Arg Leu Gln Glu Ile Glu
785 790 795 800

Glu Asn Arg Gln Glu Arg Met Ser Lys Arg Arg Gln Gln Arg Cys His
805 810 815

Gln Thr Gln Asn Phe Tyr Lys Ser Leu Gln Leu Leu Gln Ala Gly Lys
820 825 830

Leu Leu Arg Glu Gly Gly Glu Ala Gly Val Ala Glu Asp Gly Thr Pro
835 840 845

Phe Glu Asp Tyr Ser Ile Phe Leu Tyr Arg Gln Gln Ala Pro Val Phe
850 855 860

Asn Asp Arg Val Lys Asp Leu Glu Arg Lys Leu Leu Phe Pro Asp Arg
865 870 875 880

Glu Arg Gly Asp Ile Pro Ser Ala Leu Pro Arg Thr Ala Asp Glu Gln
885 890 895

Phe Ser Asp Arg Ile Lys Asn Met Glu Gln Arg Met Thr Gly Arg Gly
900 905 910

Gly Leu Gly Gly Asp Lys Lys Pro Lys Asp Leu Met Arg Ala Ile Gly
915 920 925

Lys Ile Asp Ser Asn Asp Trp Asn Val Arg Glu Ile Glu Lys Lys Ile
930 935 940

Glu Leu Ser Lys Lys Thr Glu Ile His Gly Pro Lys Gly Arg Glu Lys
945 950 955 960

Val Pro Lys Trp Ser Lys Glu Gln Phe Gln Ala Arg Gln His Lys Met
965 970 975

Ser Lys Pro Gln Arg Gln Asp Ser Arg Glu Ala Glu Lys Phe Lys Asp

980 985 990

Ile Asp Gln Thr Ile Arg Asn Leu Asp Lys Gln Leu Lys Glu Gly His
995 1000 1005

Asn Leu Asp Val Gly Glu Arg Gly Arg Asn Lys Val Ala Ser Ile
1010 1015 1020

Ala Gly Gln Phe Gly Lys Lys Asp Glu Ala Asn Ser Asp Glu Lys
1025 1030 1035

Asn Ala Gly Ser Ser Asn Ala Thr Thr Asn Thr Asn Asn Thr Val
1040 1045 1050

Ile Pro Lys Ser Ser Ser Lys Val Ala Leu Ala Phe Lys Lys Gln
1055 1060 1065

Ala Ala Ser Glu Lys Cys Arg Phe Cys Lys Gln Thr Val Tyr Pro
1070 1075 1080

Met Glu Lys Thr Thr Val Glu Gly Leu Val Leu His Arg Asn Cys
1085 1090 1095

Leu Lys Cys His His Cys His Thr Asn Leu Arg Leu Gly Gly Tyr
1100 1105 1110

Ala Phe Asp Arg Asp Asp Pro Gln Gly Arg Leu Tyr Cys Thr Gln
1115 1120 1125

His Phe Arg Leu Pro Pro Lys Pro Leu Pro Gln Arg Thr Asn Lys
1130 1135 1140

Ala Arg Lys Ser Ala Ala Ala Gln Pro Ala Ser Pro Ala Val Pro
1145 1150 1155

Pro Thr Ala Gly Ser Val Pro Thr Ala Ala Ala Thr Ser Glu His
1160 1165 1170

Met Asp Thr Thr Pro Pro Arg Asp Gln Val Asp Leu Leu Gln Thr
1175 1180 1185

Ser Arg Ala Asn Ala Ser Ala Asp Ala Met Ser Asp Asp Glu Ala
1190 1195 1200

Asn Val Ile Asp Glu His Glu Trp Ser Gly Arg Asn Phe Leu Pro
1205 1210 1215

Glu Ser Asn Asn Asp Ser Gln Ser Glu Leu Ser Ser Ser Asp Glu
1220 1225 1230

Ser Asp Thr Glu Ser Asp Ser Glu Met Phe Glu Glu Ala Asp Asp
1235 1240 1245

Ser Pro Phe Gly Ala Gln Thr Leu Gln Leu Ala Ser Asp Trp Ile
1250 1255 1260

Gly Lys Gln Tyr Cys Glu Asp Ser Asp Asp Ser Asp Asp Phe Tyr
1265 1270 1275

Asp Ser Ser Glu Asp Asp Gly Lys Asp Asp Thr Glu Gly Glu Glu
1280 1285 1290

Phe Lys Lys Ala Arg Glu Leu Arg Arg Gln Glu Val Arg Leu Gln
1295 1300 1305

Pro Leu Pro Ala Asn Leu Pro Thr Asp Thr Glu Thr Glu Lys Leu
1310 1315 1320

Lys Leu Asn Val Asp Asn Lys Glu Asn Met Ala Asp Gly Ser Ser
1325 1330 1335

Leu Lys Ser Gly Asn Ser Phe Glu Ser Ala Arg Ser Gln Pro Ser
1340 1345 1350

Thr Pro Leu Ser Thr Pro Thr Arg Val Glu Met Glu Gln Leu Glu
1355 1360 1365

Arg Asp Ala Pro Arg Lys Phe Ser Ser Glu Ile Glu Ala Ile Ser
1370 1375 1380

Glu Lys Leu Tyr His Met Asn Asn Met Val Lys Met Asn Lys Asp
1385 1390 1395

Leu Glu Val Leu Ala Lys Glu Asn Leu Val Lys Ser Gly Ile Leu
1400 1405 1410

Arg Lys Leu Thr Leu Lys Glu Lys Trp Leu Ala Glu Asn Ala Ala
1415 1420 1425

Ile Ala Ala Gly Gln Lys Val Thr Pro Thr Pro Ser Ala Thr Ala
1430 1435 1440

Pro Gly Leu Gln Pro Lys Ser Lys Phe Asp Glu Lys Phe Glu Lys
1445 1450 1455

Val Val Ser Pro Pro Gln Pro Val Val Glu Pro Lys Pro Lys Pro
1460 1465 1470

Val Ile Asp Phe Asn Leu Asp Glu Leu Lys Pro Arg Lys Pro Asn
1475 1480 1485

Phe Glu Glu Arg Pro Lys Glu Gln Leu Pro Arg Pro Glu Ser Leu
1490 1495 1500

Lys Lys Pro Pro Gln Gln Lys Pro Lys Gly Ser Ser Thr Asn Val
1505 1510 1515

Ser Arg Ser Asn Ser Leu Lys Ser Asn Ala Ser Asn Gly Ser Pro
1520 1525 1530

Lys Val Lys Lys Ala Pro Val Ser Asn Asn Ser Lys Met Gln Ile
1535 1540 1545

Glu Gly Ile Leu Asp Thr Leu Arg Lys Ile Gln Ser Gln Asn Ser
1550 1555 1560

Ser Asp Gln Asp Glu Asp Met Asp Val Asp Glu Asp Val Glu Arg
1565 1570 1575

Lys Pro Asn Lys Glu Leu Asn Ser Lys Leu Lys Glu Ile Gln Ala
1580 1585 1590

Ser Ser Phe Ala Gly Thr Met Asp His Ile Lys Ser Gln Leu Thr
1595 1600 1605

Met Pro Thr Val Ser Ala Gln Ala Pro Pro Ser Met Asp Leu Ser
1610 1615 1620

Lys Tyr Phe Pro Asn Gln Lys Gln Glu Lys Ser Ser Thr Ser Ser
1625 1630 1635

Thr Asn Lys Asn Gln Val Thr Leu Lys Asp Val Asn Leu Ala Lys
1640 1645 1650

Tyr Phe Pro Ser Ser Pro Ala Pro Gln Arg Arg Thr Val Glu Thr
1655 1660 1665

Val Ala Asp Arg Leu Lys Lys Ser Gln Thr Glu Ala Ala Leu Ala
1670 1675 1680

Lys Thr Lys Leu Leu Glu Asp Gln Ala Asn Asn Gln Ala Glu Lys
1685 1690 1695

Thr Lys Lys Glu Val Glu Lys Glu Gly Glu Ser Lys Lys Ile Thr
1700 1705 1710

Lys Lys Val Ala Asp Ser Lys Ala Val Pro Pro Lys Arg Gln Ala
1715 1720 1725

Ser Leu Asp Thr Phe Ser Leu Arg Glu His Gln Met Asp Gly Ala
1730 1735 1740

Leu Asp Leu Thr Lys Lys Lys Gly Pro Thr Lys Ala Ser Ala Gly
1745 1750 1755

Val Lys Lys Pro Ala Lys Ser Gly Ser Thr Thr Ser Val Thr Lys
1760 1765 1770

Ala Thr Ala Thr Ser Glu Gly Lys Thr Ile Lys Ile Val Lys Lys
1775 1780 1785

Ile Val Pro Lys Gly Thr Lys Ala Lys Lys Ala Ala Glu Ala Ala
1790 1795 1800

Gln Glu Ser Ala Val Val Glu Ala Pro Pro Glu Lys Lys Pro Pro
1805 1810 1815

Lys Asp Glu Ala Glu Arg Ile Leu Asp Glu Ile Leu Gly Asp Gly
1820 1825 1830

Glu Tyr Arg Ser Pro Ser Ser Glu Tyr Gln Arg Leu Phe Gln Asp
1835 1840 1845

Glu Lys Ser Pro Ser Asp Leu Ser Asp Asn Ile Asp Arg Ile Leu
1850 1855 1860

Glu Glu Ser Glu Leu Asp Val Glu Leu Gly Leu Pro Lys Arg Ser
1865 1870 1875

Ser Lys Lys Leu Val Lys Thr Lys Ser Leu Gly Glu Gly Asp Phe
1880 1885 1890

Asp Met Lys Pro Ser Lys Glu Arg Leu Thr Gly Val Gln Asn Ile

1895 1900 1905

Leu Lys Arg Phe Glu Ser Met Ser Ser Val Thr Ser Gln Asn Ser
1910 1915 1920

Asp Glu Gln Ala Gly Phe Lys Leu Arg Arg Met Glu Ser Thr Thr
1925 1930 1935

Ser Asn Leu Ser Ser Leu Thr Arg Ser Arg Glu Ser Leu Val Ser
1940 1945 1950

Val Ser Asp Ser Met Ser Asp Leu Glu Lys Thr Met Asp Tyr Leu
1955 1960 1965

Arg Asn Glu Trp Arg Asn Glu Ala Thr Asn Phe Leu Gln Lys Lys
1970 1975 1980

Arg Asp Lys Phe Tyr Ala Lys Lys Glu Glu Gln Glu Lys Glu Ala
1985 1990 1995

Lys Ile Leu Ala Lys Pro Asp Pro Leu Asp Asn Leu Pro Val Gln
2000 2005 2010

Tyr Arg Asp Ser Lys Leu Ala Lys Phe Phe Gly Leu Ala Ala Ser
2015 2020 2025

Lys Ser Pro Glu Asn Arg Lys Ser Pro Ile Lys Lys Lys Lys Ser
2030 2035 2040

Pro Ser Lys Thr Pro Lys Val Thr Lys Ala Asn Asn Ser Leu Glu
2045 2050 2055

Glu Leu Ala Lys Ile Ser Asn Val Arg Gln Thr Lys Lys Ala Gln
2060 2065 2070

Pro Lys Thr Leu Lys Pro Val Glu Val Lys Pro Leu Lys Pro Ala
2075 2080 2085

Ser Pro Val Pro Asp Asp Phe Glu Ile Leu Asp Leu Leu Glu Lys
2090 2095 2100

Ala Thr Glu Ala Lys Glu Leu Glu Arg Ser Lys Thr Lys Ser Pro
2105 2110 2115

Ala Val Glu Ser Ile Ser Gln Thr Pro Lys Glu Ala Ile Val Glu
2120 2125 2130

Ile Ser Leu Pro Val Glu Asp Ile Lys Asn Leu Pro Lys Thr Gly
2135 2140 2145

Cys Asp Lys Ser Ser Asn Ser Ser Arg Arg Gly Ser Gln Ser Ser
2150 2155 2160

Leu Ile Met Ser Arg Arg His Ser Glu Ile Ser Leu Asn Glu Lys
2165 2170 2175

Leu Asn Gln Asp Ala Leu Ala Ala Leu Asn Gln Ile Glu Lys Glu
2180 2185 2190

Arg Glu Ala Glu Gln Val Asp Glu Leu Phe Gln Ser Met Val Glu
2195 2200 2205

Glu Met Glu Gln Glu Pro Gln Pro Thr Ala Ile Val Glu Pro Pro
2210 2215 2220

Glu Glu Asp Ile Asp Ala Asp Ser Leu Cys Thr Thr Ile Ser Lys
2225 2230 2235

Ser Pro Ser Ala Gln Pro Val Thr Val Val Lys Arg Gly Ser Ser
2240 2245 2250

Glu Asp Gln Ser Ile Glu Lys Leu Phe Ser His Phe Ser Asp Glu
2255 2260 2265

Met Leu Val Asn Val Glu Phe Asp Ser Asn Asp Glu Leu Val Gly
2270 2275 2280

Ile Thr Pro Arg Ala Thr Leu Val Ser Arg Asn Thr Glu Asp Arg
2285 2290 2295

Asp Tyr Leu Asp Lys Leu Glu Ser Leu Glu Arg Asp Glu Glu Thr
2300 2305 2310

Phe Gln Pro Val Val Gly Glu Lys Phe Ile Gln Glu Asn Val Gln
2315 2320 2325

Asp Glu Val Asp Gly Leu His Phe Pro Ser Arg Pro Gln Arg Arg
2330 2335 2340

Pro Lys Ser Ser Ser Ser Ser Ser Glu Pro Ser Leu Pro Val Ala
2345 2350 2355

Pro Gln Arg Leu Glu Lys Lys Leu Ser Lys Leu Asp Pro Glu Asp
2360 2365 2370

Met Pro Pro Ser Val Gln Asp Leu Leu Gln Gln Val Tyr Gln Lys
2375 2380 2385

Asn Ile Gln Pro Glu Leu Val Glu Val Ile Pro Val Glu Gly Lys
2390 2395 2400

Gln Thr Leu Arg Phe Pro Ser Met Leu Ala Glu Glu Asp Val Asp
2405 2410 2415

Glu Val Asp His Ser Lys Glu Gly Ile Lys Lys Ile Glu Thr Ala
2420 2425 2430

Pro Glu Glu Val Arg Lys Val Thr Glu Pro Glu Asp Val Ala Arg
2435 2440 2445

Val Ile Pro Ser Pro Ile Lys Pro Ser Ile Ser Gln Ser Asn Ser
2450 2455 2460

Leu Lys Ser Glu Asn Ser Ser Gly Ser Ser Leu Val Glu Ile Pro
2465 2470 2475

Lys Ile Ile Ala Pro Pro Lys Ser Ser Ser Lys Glu Asn Ser Ser
2480 2485 2490

Asp Trp Asp Arg Glu Lys Leu Pro Ala Ser Pro Met Pro Arg Arg
2495 2500 2505

Arg Leu Leu Pro Asn Gln Thr Pro Tyr Lys Ala Pro Ser Val Ala
2510 2515 2520

Ser Lys Glu Ser Ser Leu Glu Trp Asp Met Glu Lys Leu Pro Asn
2525 2530 2535

Ser Pro Met Leu Pro Arg Arg Asn Lys Met Arg Ala Ile Ser Pro
2540 2545 2550

Ser Thr Asn Pro Val Gln Leu Leu Asn Asn Leu Pro Ser Asp Val
2555 2560 2565

Asp Asp Glu Ala Ala Gln Arg Arg Leu Ile Glu Asp Phe Glu Gln
2570 2575 2580

Glu Arg Arg Gln Ala Leu Ile Lys Arg Asp Glu Asn Phe Glu Ala
2585 2590 2595

Ile Ala Ala Glu Gln Arg Arg Arg Asp Ser Leu Gln Ser Ser Ser
2600 2605 2610

Asn Ser Ser Ser Lys Arg Ser Leu Pro Pro Pro Thr Pro Pro Met
2615 2620 2625

Met Ala Ser Arg Arg Gly Thr Thr Gln Asp Thr Asn Arg Thr Gln
2630 2635 2640

Asp Thr Ala Ser Arg His Glu Gly Thr Pro Pro Met Phe Lys Lys
2645 2650 2655

Leu Asp Val Asp Gly Ser Gly Thr Ser Met Asp Ser Thr Ser Cys
2660 2665 2670

Ser Thr Arg Arg Ser Ser Phe Ala Phe Ile Glu Leu Gln Asp Asn
2675 2680 2685

Lys Pro Val Ile Val Pro Met Pro Lys Lys Leu Lys Leu Pro Lys
2690 2695 2700

Pro Glu Pro Pro Arg Phe Val Pro Glu Pro Val Ala Thr Asp Glu
2705 2710 2715

Pro Val Pro Glu Val Phe Gln Gly Arg Ala Trp Pro Lys Thr Gln
2720 2725 2730

Leu Glu Gly Glu Val Asp Leu Gly Asp Ser Asp Asn Glu Asp Glu
2735 2740 2745

Thr Glu Lys Leu Lys Lys Gln Leu Pro Glu Tyr Ala Arg Ser Asp
2750 2755 2760

Ser Pro Pro Ser Ala Ala Phe Lys Asn Arg Lys Trp Pro Asp Gly
2765 2770 2775

Lys Thr Val Phe Asp Lys Arg Ala Glu Ser Leu Glu Glu Glu Asp
2780 2785 2790

Ile Phe Glu Gly Leu Pro Ser Pro Arg Lys Arg Gly Ser Gln Arg
2795 2800 2805

Phe Met Asp Lys Pro Arg Ser Gln Ser Pro Gln Pro Phe Lys Pro

2810 2815 2820

Leu Ala Asn Ser Ser Arg Lys Ser Ser Lys Ser Phe Ser Asp Leu
2825 2830 2835

Lys Lys Gly Pro Ser Leu Gln Ser Leu Ser Ala Gln Ser Ser Gln
2840 2845 2850

Asp Thr Asp Thr Leu Ser Thr Thr Thr Thr Val Ala Thr Ala Arg
2855 2860 2865

Pro Ala Ser Tyr Ala Asn Tyr Glu Asp Pro Met Asp Ala Ser Thr
2870 2875 2880

Gln Ala Leu Leu Asp Arg Ser Lys Arg Leu His Asn Arg Lys Arg
2885 2890 2895

Asp Phe Val Asn Glu Arg Val Val Glu Arg Asn Pro Tyr Met Arg
2900 2905 2910

Asp Val Leu Arg Ser Thr Asp Arg Arg Asp Tyr Asp Asp Val Asp
2915 2920 2925

Glu Asp Leu Thr Ser Tyr Arg Pro Arg His Tyr Ala Ser Ser Thr
2930 2935 2940

Leu Asn Arg Phe Pro Asn Thr Thr Ile Arg Lys Ser Asn Asn Tyr
2945 2950 2955

Asp Tyr Leu Ser Pro Ser Ser Asp Tyr Leu Ser Arg Arg Ser Tyr
2960 2965 2970

Ile Pro Ser Ala Ser Ala Thr Ser Ser Tyr Tyr Pro Ser Thr Thr
2975 2980 2985

Arg Ser Ser His Leu Ser Asp Leu Phe Arg Arg Arg Ser Pro Ala
2990 2995 3000

Ser Gly Thr Val Ser Ala Leu Ser Gly Tyr Gly Asn Lys Glu Ser
3005 3010 3015

Cys Val Ile Ser Ile Gly Leu Ala Leu Asp Arg Val Gly His Leu
3020 3025 3030

Ile Glu Ser Lys Cys Thr Trp Val Arg Ser Thr Lys Val Gln Thr
3035 3040 3045

Glu Ser Glu Ser Thr Ser Pro Asp Glu Val Glu Leu Asn Ser Ala
3050 3055 3060

Thr Glu Ile Ser Thr Asp Ser Glu Phe Asp Asn Asp Glu Ile Ile
3065 3070 3075

Arg Gln Ala Pro Lys Ile Phe Ile Asp Asp Thr His Leu Arg Lys
3080 3085 3090

Pro Thr Lys Val Gln Ile Lys Ser Thr Met Ile Gly Pro Asn Ala
3095 3100 3105

Ala Ser Ala Gly Leu His Gln Lys Gln Leu Ala Ala Arg Glu Lys
3110 3115 3120

Gly Gly Ser Tyr Leu Gln Lys Tyr Gln Pro Gln Pro Pro Leu Pro
3125 3130 3135

Gln Phe Arg Pro Leu Val Gln Val Asp Pro Thr Leu Leu Ile Gly
3140 3145 3150

Ser Gln Arg Ala Pro Leu Gln Asn Pro Arg Pro Gly Asp Tyr Leu
3155 3160 3165

Leu Asn Lys Thr Ala Ser Thr Glu Gly Ile Ala Ser Lys Lys Ser
3170 3175 3180

Leu Gly Leu Lys Lys Arg Tyr Leu Leu Gly Glu Pro Ala Asn Gly
3185 3190 3195

Asn Lys Ile Gln Lys Ser Gly Ser Thr Ser Val Leu Asp Ser Arg
3200 3205 3210

Ile Arg Ser Phe Gln Ser Asn Ile Ser Glu Cys Gln Lys Leu Leu
3215 3220 3225

Asn Pro Ser Ser Asp Ile Ser Ala Gly Met Arg Thr Phe Leu Asp
3230 3235 3240

Arg Thr Lys Leu Gly Glu Gly Ser Gln Thr Thr Pro Gly Gln Thr
3245 3250 3255

Asn Glu Leu Ile Arg Ser Ala Thr Ser Asn Val Ile Asn Asp Leu
3260 3265 3270

Arg Val Glu Leu Arg Ile Gln Lys Thr Gly Ser Ser His Ser Thr
3275 3280 3285

Asp Asn Glu Lys Glu Asn Val Phe Val Asn Cys Lys Asn Glu Leu
3290 3295 3300

Asn Lys Gly Met Glu Tyr Thr Asp Ala Val Asn Ala Thr Leu Leu
3305 3310 3315

Asp Gln Leu Ala Arg Lys Ser Ser Pro Thr Thr Pro Thr Asn Lys
3320 3325 3330

Thr Val Val Glu Val Ile Asp Leu Val Thr Pro Glu Lys Pro Ile
3335 3340 3345

Asp Ile Ile Asp Leu Thr Ala Leu Glu Thr Pro Lys Lys Gln Leu
3350 3355 3360

Val Asp Gly Ser Ala Met Asp Val Asp Glu Arg Leu Thr Pro Asp
3365 3370 3375

Ser Asn Lys Ile Ser Glu Leu Gln Gln Glu Val Lys Glu Glu Pro
3380 3385 3390

Lys Pro Asp Val Ser Arg Asp Val Lys Glu Cys Ile Pro Asp Ile
3395 3400 3405

Leu Gly His Ile Lys Glu Gly Thr Gly Ser Lys Glu Pro Gly Gly
3410 3415 3420

Glu Asp Gln Gln Ser Leu Leu Glu Gln Ser Asp Glu Glu Lys Arg
3425 3430 3435

Asp Ser Pro Glu Lys Asp Val Ala Glu His Glu Leu Tyr Glu Pro
3440 3445 3450

Asp Ser Val Gln Ile Gln Val Pro Asn Ile Pro Trp Glu Lys Ser
3455 3460 3465

Lys Pro Glu Val Met Ser Thr Thr Gly Ser Ser Gly Ser Ile Cys
3470 3475 3480

Ser Ser Ser Asp Ser Ser Ser Ile Glu Asp Ile Gln His Tyr Ile
3485 3490 3495

Leu Glu Ser Thr Thr Ser Pro Asp Thr Gln Thr Val Gly Gly Lys
3500 3505 3510

His Asn Val Pro Arg Leu Glu Val His Asp Thr Ser Gly Ala Leu
3515 3520 3525

Met Gln Val Asp Ser Leu Met Ile Val Asn Gly Lys Tyr Ile Gly
3530 3535 3540

Asp Pro Glu Asp Val Lys Phe Leu Asp Met Pro Ala Asn Val Ile
3545 3550 3555

Val Pro Pro Ala Pro Ala Leu Lys Thr Asn Glu Leu Asp Met Glu
3560 3565 3570

Asp Asp Gln Glu Ala Glu Ala Glu Pro Val Thr Ala Thr Pro Glu
3575 3580 3585

Pro Val Glu Cys Thr Val Ile Glu Ala Glu Arg Arg Val Thr Ala
3590 3595 3600

Pro Pro Pro Leu Pro Glu Met Gly Pro Pro Lys Leu Lys Phe Asp
3605 3610 3615

Ser Lys Asn Glu Asn Lys Ile Glu Ser Leu Lys Asn Leu Pro Leu
3620 3625 3630

Ile Val Glu Ser Asn Val Glu His Ser Gln Ala Val Lys Pro Ile
3635 3640 3645

Thr Leu Asn Leu Ser Asn Leu Ala Arg Thr Pro Asp Thr Pro Thr
3650 3655 3660

Thr Pro Thr Ala His Asp Ser Asp Lys Thr Pro Thr Gly Glu Ile
3665 3670 3675

Leu Ser Arg Gly Ser Asp Ser Glu Thr Glu His Thr Gly Thr Gly
3680 3685 3690

Gln Val Leu Thr Glu Thr Glu Leu Ser Asp Trp Thr Ala Asp Asp
3695 3700 3705

Cys Ile Ser Glu Asn Phe Val Asp Leu Glu Phe Ala Leu Asn Ser
3710 3715 3720

Asn Lys Gly Thr Ile Lys Arg Arg Lys Asp Arg Arg Arg Ser Gly

3725 3730 3735

Ala Ser Lys Leu Pro Ser Gly Asn Glu Val Ile His Glu Leu Ala
3740 3745 3750

Arg Gln Ala Pro Val Val Gln Met Asp Gly Ile Leu Ser Ala Ile
3755 3760 3765

Asp Ile Asp Asp Ile Glu Phe Met Asp Thr Gly Ser Glu Gly Ser
3770 3775 3780

Cys Ala Glu Ala Tyr Pro Ala Thr Asn Thr Ala Leu Ile Gln Asn
3785 3790 3795

Arg Gly Tyr Met Glu Tyr Ile Glu Ala Glu Pro Lys Lys Thr Thr
3800 3805 3810

Arg Lys Ala Ala Pro Pro Ser Ser Tyr Pro Gly Asn Leu Pro Pro
3815 3820 3825

Leu Met Thr Lys Arg Asp Glu Lys Leu Gly Val Asp Tyr Ile Glu
3830 3835 3840

Gln Gly Ala Tyr Ile Met His Asp Asp Ala Lys Thr Pro Val Asn
3845 3850 3855

Glu Val Ala Pro Ala Met Thr Gln Ser Leu Thr Asp Ser Ile Thr
3860 3865 3870

Leu Asn Glu Leu Asp Asp Asp Ser Met Ile Ile Ser Gln Thr Gln
3875 3880 3885

Pro Thr Thr Thr Glu Glu Ser Glu Ala Leu Thr Val Val Thr Ser
3890 3895 3900

Pro Leu Asp Thr Ser Ser Pro Arg Val Leu Asp Gln Phe Ala Ser
3905 3910 3915

Met Leu Ala Ala Gly Lys Gly Asp Ser Thr Pro Ser Ser Ser Glu
3920 3925 3930

Gln Gln Pro Lys Thr Ser Thr Val Thr Ser Ser Ser Thr Gly Pro
3935 3940 3945

Asn Ser Ser Thr Thr Gly Asn Val Ser Lys Glu Pro Gln Glu Glu
3950 3955 3960

Asp Leu Gln Ile Gln Phe Glu Tyr Val Arg Ala Leu Gln Gln Arg
3965 3970 3975

Ile Ser Gln Ile Ser Thr Gln Arg Arg Lys Ser Ser Lys Gly Glu
3980 3985 3990

Ala Pro Asn Leu Gln Leu Asn Ser Ser Ala Pro Val Ile Glu Ser
3995 4000 4005

Ala Glu Asp Pro Ala Lys Pro Ala Glu Glu Pro Leu Val Ser Met
4010 4015 4020

Arg Pro Arg Thr Thr Ser Ile Ser Gly Lys Val Pro Glu Ile Pro
4025 4030 4035

Thr Leu Ser Ser Lys Leu Glu Glu Ile Thr Lys Glu Arg Thr Lys
4040 4045 4050

Gln Lys Asp Leu Ile His Asp Leu Val Met Asp Lys Leu Gln Ser
4055 4060 4065

Lys Lys Gln Leu Asn Ala Glu Lys Arg Leu His Arg Ser Arg Gln
4070 4075 4080

Arg Ser Leu Leu Thr Ser Gly Tyr Ala Ser Gly Ser Ser Leu Ser
4085 4090 4095

Pro Thr Pro Lys Leu Ala Ala Ala Cys Ser Pro Gln Asp Ser Asn
4100 4105 4110

Cys Ser Ser Gln Ala His Tyr His Ala Ser Thr Ala Glu Glu Ala
4115 4120 4125

Pro Lys Pro Pro Ala Glu Arg Pro Leu Gln Lys Ser Ala Thr Ser
4130 4135 4140

Thr Tyr Val Ser Pro Tyr Arg Thr Val Gln Ala Pro Thr Arg Ser
4145 4150 4155

Ala Asp Leu Tyr Lys Pro Arg Pro Phe Ser Glu His Ile Asp Ser
4160 4165 4170

Asn Ala Leu Ala Gly Tyr Lys Leu Gly Lys Thr Ala Ser Phe Asn
4175 4180 4185

Gly Gly Lys Leu Gly Asp Phe Ala Lys Pro Ile Ala Pro Ala Arg
4190 4195 4200

Val Asn Arg Gly Gly Gly Val Ala Thr Ala Asp Ile Ala Asn Ile
4205 4210 4215

Ser Ala Ser Thr Glu Asn Leu Arg Ser Glu Ala Arg Ala Arg Ala
4220 4225 4230

Arg Leu Lys Ser Asn Thr Glu Leu Gly Leu Ser Pro Glu Glu Lys
4235 4240 4245

Met Gln Leu Ile Arg Ser Arg Leu His Tyr Asp Gln Asn Arg Ser
4250 4255 4260

Leu Lys Pro Lys Gln Leu Glu Glu Met Pro Ser Gly Asp Leu Ala
4265 4270 4275

Ala Arg Ala Arg Lys Met Ser Ala Ser Lys Ser Val Asn Asp Leu
4280 4285 4290

Ala Tyr Met Val Gly Gln Gln Gln Gln Gln Gln Val Glu Lys Asp
4295 4300 4305

Ala Val Leu Gln Ala Lys Ala Ala Asp Phe Thr Ser Asp Pro Asn
4310 4315 4320

Leu Ala Ser Gly Gly Gln Glu Lys Ala Gly Lys Thr Lys Ser Gly
4325 4330 4335

Arg Arg Pro Lys Asp Pro Glu Arg Arg Lys Ser Leu Ile Gln Ser
4340 4345 4350

Leu Ser Ser Phe Phe Gln Lys Gly Ser Gly Ser Ala Ala Ser Ser
4355 4360 4365

Ser Lys Glu Gln Gly Gly Ala Val Ala Ala Val His Ser Glu Gln
4370 4375 4380

Ser Glu Arg Pro Gly Thr Ser Ser Ser Gly Thr Pro Thr Ile Ser
4385 4390 4395

Asp Ala Ala Gly Gly Gly Gly Gly Gly Gly Gly Val Phe Ser Arg
4400 4405 4410

Phe Arg Ile Ser Pro Lys Ser Lys Glu Lys Ser Lys Ser Cys Phe
4415 4420 4425

Asp Leu Arg Asn Phe Gly Phe Gly Asp Lys Asp Met Leu Val Cys
4430 4435 4440

Asn Ala Ala Ser Pro Ala Gly Ala Thr Ser Ala Ser Gln Lys Asn
4445 4450 4455

His Ser Gln Glu Tyr Leu Asn Thr Thr Asn Asn Ser Arg Tyr Arg
4460 4465 4470

Lys Gln Thr Asn Thr Ala Lys Pro Lys Pro Glu Ser Phe Ser Ser
4475 4480 4485

Ser Ser Pro Gln Leu Tyr Ile His Lys Pro His His Leu Ala Ala
4490 4495 4500

Ala His Pro Ser Ala Leu Asp Asp Gln Thr Pro Pro Pro Ile Pro
4505 4510 4515

Pro Leu Pro Leu Asn Tyr Gln Arg Ser Asp Asp Glu Ser Tyr Ala
4520 4525 4530

Asn Glu Thr Arg Glu His Lys Lys Gln Arg Ala Ile Ser Lys Ala
4535 4540 4545

Ser Arg Gln Ala Glu Leu Lys Arg Leu Arg Ile Ala Gln Glu Ile
4550 4555 4560

Gln Arg Glu Gln Glu Glu Ile Glu Val Gln Leu Lys Asp Leu Glu
4565 4570 4575

Ala Arg Gly Val Leu Ile Glu Lys Ala Leu Arg Gly Glu Ala Gln
4580 4585 4590

Asn Ile Glu Asn Leu Asp Ala Thr Lys Asp Asn Asp Glu Lys Leu
4595 4600 4605

Leu Lys Glu Leu Leu Glu Ile Trp Arg Asn Ile Thr Ala Leu Lys
4610 4615 4620

Lys Arg Asp Glu Glu Leu Thr Ile Arg Gln Gln Glu Leu Gln Leu
4625 4630 4635

Glu Tyr Arg His Ala Gln Leu Lys Glu Glu Leu Asn Leu Arg Leu

4640 4645 4650

Ser Cys Asn Lys Leu Asp Lys Ser Ser Ala Asp Val Ala Ala Glu
4655 4660 4665

Gly Ala Ile Leu Asn Glu Met Leu Glu Ile Val Ala Lys Arg Ala
4670 4675 4680

Ala Leu Arg Pro Thr Ala Ser Gln Leu Asp Leu Thr Ala Ala Gly
4685 4690 4695

Ser Ala Ser Thr Ser Ala Glu Ala Thr Gly Ile Lys Leu Thr Gly
4700 4705 4710

Gln Pro His Asp His Glu Glu Ser Ile Ile
4715 4720

<210> 9
<211> 9009
<212> DNA
<213> Drosophila

<400> 9
atgagccgcc aacaccagcg gcaccaccag cagcatcacc acctgccgcc gcaccagcaa 60

ccgcagcagc agatgccgca acaacagcag cagctgacgg cgcagcagca gcaacaacag 120

cagctgctga tggcggagca cgcggcggcc gcggaggcgg cggagctatt cgacctgctg 180

tgcgtggcca caacgatgcg ccagatcctg gcgctccatc gggccatgtg cgaggctgtg 240

ggattgagac cctcgcctct gaacgacttc tacccacggc taaaggccaa ggtgcgttcg 300

tggaaggcgc aggccctgtg gaagaagttc gacgccagag ctgcccatag agtctacggc 360

aagggagctg cctgtactgg cacacgcgtc ctggtcatcg gagcagggcc ctgtggactg 420

cgcaccgcca tcgaggccca actgctgggc gccaaggtgg tggtgctgga gaaacgcgat 480

cgcatcaccc ggaacaatgt gctccatctg tggccattcg tcatcacgga tctgcgcaac 540

ttgggcgcaa agaagttcta cggcaagttt tgcgccggct ccatcgatca catctccatt 600

cggcagctgc agtgcatgct gctcaaggtg gcgctgctcc tgggcgtaga gatccacgag 660

ggagtcagtt ttgatcacgc tgtagagccc tctggcgatg gcggcggatg gagggcagct 720

gttactcccg cagatcatcc tgtatctcac tacgaattcg atgtgttgat cggagcggat 780

ggcaagcgga atatgctgga ctttaggagg aaggagttcc gcgggaagct ggccatcgct 840

attacagcga actttatcaa caagaagacg gaggcggagg ctaaagtaga ggagatcagt 900

ggggtggctt tcatcttcaa ccaggccttc ttcaaggagc tgtacgggaa gacgggcatc 960

gacctggaaa acatcgtcta ctacaaggac gagacgcact acttcgtgat gacggccaag 1020
aagcacagtc taattgacaa gggcgttatt atcgaggata tggccgatcc cggcgagctt 1080
ctcgccccag ccaatgtgga tacacaaaag ctgcacgact atgcacgcga ggctgcggag 1140
ttctccaccc aataccaaat gccaaacctg gagttcgctg ttaatcacta cggcaaacca 1200
gatgtggcca tgttcgactt cacatcgatg tttgccgccg agatgtcctg tcgggtgatt 1260
gtgcgcaaag gagctcgcct gatgcagtgc ctcgtgggtg acagtctgct cgagccgttt 1320
tggcccactg gatcgggttg tgcccgtgga ttcttatcca gcatggatgc tgcctatgcc 1380
atcaagcttt ggtccaaccc gcagaacagc acacttggcg ttctggcgca gcgcgaaagc 1440
atctaccggc tgcttaacca gaccacgccg gacaccctgc agcgggacat cagtgcctat 1500
accgtggatc cggccacgcg ctatccgaat ctgaacaggg agtcggtcaa tagctggcag 1560
gtcaaacatc tggtcgacac ggacgacccg tccattctgg agcagacctt catggacacg 1620
catgctctgc agaccccgca tttggacaca ccgggcagac gcaagcgacg cagtggagac 1680
ttgctgcccc agggtgccac gttgctgaga tggataagtg cccagctgca ttcctatcag 1740
tttattcccg aactcaagga ggcttcggat gtgttccgga atggacgcgt tctgtgtgcg 1800
cttatcaatc gctatcgtcc tgatctcatc gactacgctg ccaccaagga catgagtccc 1860
gtggagtgca atgagctgtc attcgccgtc ctagagcgcg aactccacat cgatcgcgtc 1920
atgagtgcca aacagtcgct ggacttgacc gagctggagt cgcgaatctg gctcaactat 1980
ttggaccaga tctgcgactt gtttcgcggc gagatccccc atatcaagca ccccaagatg 2040
gactttagcg atttgcgcca gaagtatcgt atcaaccata cgcatgccca acccgacttc 2100
tccaagctgc tggcaacgaa acccaaggcc aagtcgccga tgcaggatgc tgtggacata 2160
cccacgacag tgcagcggcg ctcggtgctc gaggaggagc gagccaagcg gcagcgtcgc 2220
cacgagcagc ttcttaacat cggtggaggg gcagcaggag ccgccgccgg agttgccgga 2280
agcgggacag gaaccacaac gcagggtcaa aacgatacgc cacgccggtc caagaagcgc 2340
cgtcaggttg acaaaaccgc caatattgag gagcgccagc agcgcttgca ggagatcgag 2400
gagaatcggc aggagcggat gagcaagcgg cgccagcagc gctgtcacca gacgcagaat 2460
ttctacaaga gccttcagct cctgcaggcg ggcaagctct tgagggaggg tggtgaggcg 2520
ggagtggccg aggatggcac cccattcgag gactactcga tattcctcta ccgccagcag 2580
gcgcccgtat tcaatgatcg cgtcaaggac ctcgagcgaa agcttctgtt tcccgatcgc 2640
gaacggggag atattccgtc ggcattgccg cgcacggcgg acgagcagtt cagcgatcgg 2700
attaaaaaca tggagcagcg gatgacggga cgtggtggcc ttggtggtga caagaagccc 2760
aaggatctga tgcgggctat cggcaagatc gactcgaacg attggaatgt acgcgagatc 2820

gagaagaaga tcgagctgtc gaagaagacg gagatccacg ggcctaaggg ccgcgagaag 2880
gttcccaagt ggagcaagga gcagttccag gcgcgacagc acaagatgtc caagccgcaa 2940
cgccaggatt cgcgtgaggc ggaaaagttc aaggacatcg accagactat ccgcaacttg 3000
gacaagcagc taaaggaggg ccacaatctg gatgtgggcg agcggggacg caacaaagtg 3060
gcctccattg ccggtcagtt tggcaaaaag gatgaggcca attcggatga gaagaacgcg 3120
ggcagcagca atgccaccac caacaccaac aacacagtca tacccaaatc tagttccaag 3180
gtggcgctgg cctttaaaaa gcaggctgcc tccgaaaagt gccgcttctg taagcaaacc 3240
gtctacctga tggagaagac caccgtggag ggattggttc tgcatcgcaa ttgccttaag 3300
tgccaccact gccacaccaa cttgcgtctg ggaggctacg cctttgatcg ggacgatccg 3360
cagggccgat tttactgcac ccaacacttc cggttgccac ccaaaccgct gccgcagcgc 3420
accaacaaag ccaggaaatc cgctgccgct caacccgcct cgcctgctgt accaccaact 3480
gcgggatccg tacccactgc agctgccaca tcggagcata tggacaccac tccacccagg 3540
gaccaggtgg acctactgga gacctcgcga gcaaatgcct ctgccgatgc catgtccgat 3600
gatgaagcca atgttatcga tgagcacgaa tggtctggtc gcaacttctt gcccgagtcc 3660
aacaacgatt cccaatcaga gctatccagt tcagatgagt cggatacgga atcggattcg 3720
gagatgtttg aggaggcgga tgattcgccg tttggtgctc agaccctcca gctggcgtcg 3780
gattggattg gaaagcaata ctgtgaggac agtgatgatt ctgacgattt ctacgactca 3840
agtgaaggta ttgcggatga cggcaaagat gacaccgagg gtgaggaatt caagaaggcc 3900
cgcgaattga ggcgccagga agttcgcctg cagccgttgc ccgccaatct gcccacagat 3960
acggagaccg aggttcaaac cgagtccgag agcacttcac ccgacgaagt ggagctcaat 4020
tctgccactg agatatccac cgactctgag tttgacaacg atgagattat acgccaggcg 4080
cccaaaatct tcatcgatga cacccatcta aggaagccca ccaaggttca gatcaagtcc 4140
accatgatcg gacccaatgc agcttccgcc ggactccatc agaagcagtt ggcggcgcgt 4200
gagaagggcg gcagctacct ccagaagtac caaccacaac cgccactgtc acagtttaaa 4260
ccgttggtcc aggtggatcc caccctgctc attggcagcc agcgcgctcc tcttcagaat 4320
ccacggccag gagactactt gctaaacaag acggccagta cggagggtat cgcctcaaaa 4380
aagagcctgg agctaaaaaa gcgctatctg ctgggtgagc cggccaatgg cgataagatc 4440
cagaagtccg gatccacttc agtgctggat tcacgcattc gcagcttcca gtcgaacata 4500
tcggagtgcc agaagctttt gaatcccagc agcgacataa gtgccggcat gcgaaccttc 4560
ctcgatcgca caaagttggg cgaaggcagc cagacgacac ccggacagac gaacgaacta 4620

atccgttccg ccaccagcaa tgtgattaac gatctgcgcg tggagcttcg gatacagaaa 4680
actgactcca gccactccac ggacaacgag aaggaaaacg ttttcgtgaa ctgtaagaac 4740
gagctgaaca aggggatgga atacacggat gcggtcaatg ccacgctgct ggaccagctg 4800
gccagaaaaa gttcacccac cacgccgacg aataagacgg tggtcgaggt tattgacctg 4860
gttacacctg agaagccaat tgacattatc gatctaacgg cactggaaac gccgaaaaag 4920
cagttggtcg atggtagcgc catggatgta gatgaacgcc tcacacccga tagcaacaaa 4980
atcagcgaac tgcagcagga agtgaaggag gaacccaagc cggatgtctc tagggatgtg 5040
aaagaatgca taccagatat actgggacac attaaggagg gaacgggatc gaaggagcca 5100
ggtggagagg accaacagag cctgctggag cagtcggacg aagagaagcg cgactcaccg 5160
gaaaaggatg tggccgaaca tgagctttat gaaccggaca gtgtgcagat ccaggtgccc 5220
aatatcccat gggaaaaaag caagccggag gtcatgtcta ccaccggcag cagtggctcc 5280
atctgctcaa gctcagactc ttctagtatt gaagacatcc agcactacat tttggagtcc 5340
acaactagtc cagatactca gacagttggc ggaaagcaca atgtgccccg tttggaggtg 5400
cacgacacaa gtggtgccct gatgcaggtg gacagcctga tgattgtgaa cggaaagtat 5460
attggggatc ccgaggatgt caagttcttg gatatgccgg ccaatgttat tgttccgcca 5520
gcaccggcgc ttaaaacgaa tgagctggat atggaggatg accaagaggc ggaggcggaa 5580
ccagtaactg ctactccgga gccggtggaa tgtacggtca tcgaggctga gcgccgtgtt 5640
actgctcccc ctcctttgcc agagatgggt ccacccaaac tgaagttcga tagcaaaaat 5700
gagaacaaga tcgagagctt gaagaatctt ccgttgatcg tagagagcaa tgtggagcac 5760
agtcaggcag tgaaacccat tactcttaac ttaagcaatc tggccaggac gccggataca 5820
ccaaccacgc ccacggcgca cgatagcgat aaaacaccca ctggggaaat tctctcgcga 5880
ggatctgact cagaaaccga gcacactggc actggtcagg tactaacgga gacggaactc 5940
tccgactgga cggccgacga ctgtatctcg gagaactttg ttgacttgga gttcgcgctt 6000
aactctaaca agggtacgat aaaacggcgc aaggatcgac gacgcagtgg agcaagcaaa 6060
cttcccagtg gcaacgaggt aatccacgag ctggccaggc aggcgccagt ggtgcaaatg 6120
gatggaattc ttagtgccat cgacattgat gacatagagt tcatggacac gggttcggag 6180
ggttcttgtg ctgaagctta tcccgcaaca aatacagctc tcattcagaa tagaggttac 6240
atggagtaca tcgaggcgga gccgaaaaag acgacccgca aggcagctcc accatccagt 6300
tacccaggaa atttaccgcc tttaatgacg aagcgggacg agaaactggg cgttgattac 6360
attgagcagg gggcgtacat aatgcacgat gatgcaaaga cgcctgtgaa tgaggtggct 6420
cctgccatga cccagtcgct aactgactca atcacgctca atgaactgga tgatgacagt 6480

atgataatat cccaaaccca gccaacgaca acggaggaaa gtgaggcact gacggtggtc 6540
accagtccac ttgacacgtc ctcgcccagg gttctcgatc aatttgcatc catgttggcg 6600
gcgggaaaag gtgactccac acccagtagc tcagagcaac aaccaaagac gtctacggtg 6660
acgagcagca gcactgggcc caactcctcg acaacaggaa acgtctcgaa ggagccgcag 6720
gaggaggacc tgcaaatcca gtttgagtat gttcgagcac tgcagcagcg gatatcgcag 6780
atcagcaccc aacggcgtaa gagctctaag ggagaggcac ctaacctgca gctaaacagt 6840
agcgcacctg tgatagaatc agccgaggat ccggccaagc ccgcagagga gcctctggtc 6900
tcaatgcgac cgcggaccac cagcatttcc ggaaaggtac cggagatacc cacacttagc 6960
agcaagctgg aagagataac caaagaacgc actaagcaaa aggatctgat tcacgaccta 7020
gtcatggaca agttgcagtc gaagaagcag ctaaacgctg agaagcgtct gcaccggagt 7080
cgacagcgca gtttgctgac cagtggctat gccagtggct ccagccttag tccgacgccc 7140
aagctggctg ctgcttgcag tccgcaggat tccaactgct ctagccaagc gcactaccac 7200
gcctccacgg cggaggaggc cccgaagccg ccggcggaaa ggccgttgca gaagtccgcc 7260
acgtccacct atgtgtcgcc ttatcgcact gtccaagcgc ccacacgtag tgctgatctc 7320
tataagccgc gccccttcag cgaacacatc gattcgaacg ctctggcggg ttacaagctc 7380
ggcaagacgg cctcgtttaa tggcggcaag ttgggcgact ttgcgaaacc cattgccccg 7440
gcgagagtta accgaggagg aggtgtcgcg accgcggata tagccaatat ttccgcgtcg 7500
acggagaacc taagaagcga ggccagggcc agggctcgtc ttaagtctaa cacagagctg 7560
ggccttagtc ccgaggaaaa gatgcagcta atacgttcaa gattgcacta cgaccaaaac 7620
agatctctga agccgaagca actggaggag atgccatccg gggatctggc ggcacgtgcc 7680
cgcaaaatga gtgcctcgaa gagcgtcaat gatctggcct acatggtggg acagcagcag 7740
cagcagcagg ttgagaagga tgccgtgctc caagccaagg cggctgactt tacatccgat 7800
cccaatttgg cgtccggtgg tcaggagaag gcaggcaaaa ctaagtccgg acgcaggcca 7860
aaggatccgg agcggcgtaa gagtctcata cagtcgttgt ccagcttctt ccaaaaggga 7920
tctggatccg cggcctccag ttccaaggag cagggcggcg ctgtggctgc cgtccactct 7980
gaacagtcag agcgaccagg caccagcagc agcggcacgc ccacaatatc ggatgcggcg 8040
ggtggaggcg gaggaggagg tggcgtcttc agcagattcc gcatctcgcc caagtccaag 8100
gagaagtcaa agtcttgctt tgatctgagg aatttcggtt ttggtgacaa ggatatgctg 8160
gtctgcaatg cagcatctcc agcaggagcc acatccgcat cacagaaaaa tcactcgcaa 8220
gagtatctga acaccacgaa caacagtcgc tatcgaaagc aaacgaacac tgcgaaaccg 8280

aaacccgaat cgttctcttc atccagtccg cagctctata tacacaagcc ccaccacctg 8340
gccgcagctc atcccagtgc cctggacgac cagacaccac cacccatacc gcctcttcca 8400
ctgaattatc agagatccga tgatgagagc tacgctaacg agacacgaga gcataagaag 8460
caacgtgcca tatcgaaggc ttcacgacaa gctgagctca agcgattgcg aatcgctcaa 8520
gagattcagc gggaacagga ggagatcgag gtgcaactga aggatctgga ggcacgcggc 8580
gtgcttattg agaaggcctt gcgaggcgag gcgcagaata ttgaaaacct ggatgcgaca 8640
aaggacaacg acgagaagct acttaaggaa cttttggaga tttggcgcaa catcacagca 8700
ctcaagaaac gcgatgagga actgactata aggcaacagg aactgcaact ggagtatcgg 8760
catgcccagc tgaaggaaga gctcaatctg cgcttgtcct gcaacaaact ggacaaaagc 8820
tctgccgatg tggccgccga gggagcaatt ctcaacgaga tgctggaaat tgtcgccaag 8880
cgagccgccc tacgacccac agcctcccag ctcgacctca cggcagcggg atcagcatcc 8940
acgtccgccg aggcaacggg cattaagctg acgggacaac cgcatgacca cgaagaatcg 9000
atcatttga 9009

<210> 10
<211> 3002
<212> PRT
<213> Drosophila

<400> 10

Met Ser Arg Gln His Gln Arg His His Gln Gln His His His Leu Pro
1 5 10 15

Pro His Gln Gln Pro Gln Gln Gln Met Pro Gln Gln Gln Gln Gln Leu
20 25 30

Thr Ala Gln Gln Gln Gln Gln Gln Gln Leu Leu Met Ala Glu His Ala
35 40 45

Ala Ala Ala Glu Ala Ala Glu Leu Phe Asp Leu Leu Cys Val Ala Thr
50 55 60

Thr Met Arg Gln Ile Leu Ala Leu His Arg Ala Met Cys Glu Ala Val
65 70 75 80

Gly Leu Arg Pro Ser Pro Leu Asn Asp Phe Tyr Pro Arg Leu Lys Ala
85 90 95

Lys Val Arg Ser Trp Lys Ala Gln Ala Leu Trp Lys Lys Phe Asp Ala
100 105 110

Arg Ala Ala His Arg Val Tyr Gly Lys Gly Ala Ala Cys Thr Gly Thr
115 120 125

Arg Val Leu Val Ile Gly Ala Gly Pro Cys Gly Leu Arg Thr Ala Ile
130 135 140

Glu Ala Gln Leu Leu Gly Ala Lys Val Val Val Leu Glu Lys Arg Asp
145 150 155 160

Arg Ile Thr Arg Asn Asn Val Leu His Leu Trp Pro Phe Val Ile Thr
165 170 175

Asp Leu Arg Asn Leu Gly Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala
180 185 190

Gly Ser Ile Asp His Ile Ser Ile Arg Gln Leu Gln Cys Met Leu Leu
195 200 205

Lys Val Ala Leu Leu Leu Gly Val Glu Ile His Glu Gly Val Ser Phe
210 215 220

Asp His Ala Val Glu Pro Ser Gly Asp Gly Gly Gly Trp Arg Ala Ala
225 230 235 240

Val Thr Pro Ala Asp His Pro Val Ser His Tyr Glu Phe Asp Val Leu
245 250 255

Ile Gly Ala Asp Gly Lys Arg Asn Met Leu Asp Phe Arg Arg Lys Glu
260 265 270

Phe Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Lys
275 280 285

Lys Thr Glu Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe
290 295 300

Ile Phe Asn Gln Ala Phe Phe Lys Glu Leu Tyr Gly Lys Thr Gly Ile
305 310 315 320

Asp Leu Glu Asn Ile Val Tyr Tyr Lys Asp Glu Thr His Tyr Phe Val
325 330 335

Met Thr Ala Lys Lys His Ser Leu Ile Asp Lys Gly Val Ile Ile Glu
340 345 350

Asp Met Ala Asp Pro Gly Glu Leu Leu Ala Pro Ala Asn Val Asp Thr
355 360 365

Gln Lys Leu His Asp Tyr Ala Arg Glu Ala Ala Glu Phe Ser Thr Gln
370 375 380

Tyr Gln Met Pro Asn Leu Glu Phe Ala Val Asn His Tyr Gly Lys Pro
385 390 395 400

Asp Val Ala Met Phe Asp Phe Thr Ser Met Phe Ala Ala Glu Met Ser
405 410 415

Cys Arg Val Ile Val Arg Lys Gly Ala Arg Leu Met Gln Cys Leu Val
420 425 430

Gly Asp Ser Leu Leu Glu Pro Phe Trp Pro Thr Gly Ser Gly Cys Ala
435 440 445

Arg Gly Phe Leu Ser Ser Met Asp Ala Ala Tyr Ala Ile Lys Leu Trp
450 455 460

Ser Asn Pro Gln Asn Ser Thr Leu Gly Val Leu Ala Gln Arg Glu Ser
465 470 475 480

Ile Tyr Arg Leu Leu Asn Gln Thr Thr Pro Asp Thr Leu Gln Arg Asp
485 490 495

Ile Ser Ala Tyr Thr Val Asp Pro Ala Thr Arg Tyr Pro Asn Leu Asn
500 505 510

Arg Glu Ser Val Asn Ser Trp Gln Val Lys His Leu Val Asp Thr Asp
515 520 525

Asp Pro Ser Ile Leu Glu Gln Thr Phe Met Asp Thr His Ala Leu Gln
530 535 540

Thr Pro His Leu Asp Thr Pro Gly Arg Arg Lys Arg Arg Ser Gly Asp
545 550 555 560

Leu Leu Pro Gln Gly Ala Thr Leu Leu Arg Trp Ile Ser Ala Gln Leu
565 570 575

His Ser Tyr Gln Phe Ile Pro Glu Leu Lys Glu Ala Ser Asp Val Phe
580 585 590

Arg Asn Gly Arg Val Leu Cys Ala Leu Ile Asn Arg Tyr Arg Pro Asp

595 600 605

Leu Ile Asp Tyr Ala Ala Thr Lys Asp Met Ser Pro Val Glu Cys Asn
610 615 620

Glu Leu Ser Phe Ala Val Leu Glu Arg Glu Leu His Ile Asp Arg Val
625 630 635 640

Met Ser Ala Lys Gln Ser Leu Asp Leu Thr Glu Leu Glu Ser Arg Ile
645 650 655

Trp Leu Asn Tyr Leu Asp Gln Ile Cys Asp Leu Phe Arg Gly Glu Ile
660 665 670

Pro His Ile Lys His Pro Lys Met Asp Phe Ser Asp Leu Arg Gln Lys
675 680 685

Tyr Arg Ile Asn His Thr His Ala Gln Pro Asp Phe Ser Lys Leu Leu
690 695 700

Ala Thr Lys Pro Lys Ala Lys Ser Pro Met Gln Asp Ala Val Asp Ile
705 710 715 720

Pro Thr Thr Val Gln Arg Arg Ser Val Leu Glu Glu Glu Arg Ala Lys
725 730 735

Arg Gln Arg Arg His Glu Gln Leu Leu Asn Ile Gly Gly Gly Ala Ala
740 745 750

Gly Ala Ala Ala Gly Val Ala Gly Ser Gly Thr Gly Thr Thr Thr Gln
755 760 765

Gly Gln Asn Asp Thr Pro Arg Arg Ser Lys Lys Arg Arg Gln Val Asp
770 775 780

Lys Thr Ala Asn Ile Glu Glu Arg Gln Gln Arg Leu Gln Glu Ile Glu
785 790 795 800

Glu Asn Arg Gln Glu Arg Met Ser Lys Arg Arg Gln Gln Arg Cys His
805 810 815

Gln Thr Gln Asn Phe Tyr Lys Ser Leu Gln Leu Leu Gln Ala Gly Lys
820 825 830

Leu Leu Arg Glu Gly Gly Glu Ala Gly Val Ala Glu Asp Gly Thr Pro
835 840 845

Phe Glu Asp Tyr Ser Ile Phe Leu Tyr Arg Gln Gln Ala Pro Val Phe
850 855 860

Asn Asp Arg Val Lys Asp Leu Glu Arg Lys Leu Leu Phe Pro Asp Arg
865 870 875 880

Glu Arg Gly Asp Ile Pro Ser Ala Leu Pro Arg Thr Ala Asp Glu Gln
885 890 895

Phe Ser Asp Arg Ile Lys Asn Met Glu Gln Arg Met Thr Gly Arg Gly
900 905 910

Gly Leu Gly Gly Asp Lys Lys Pro Lys Asp Leu Met Arg Ala Ile Gly
915 920 925

Lys Ile Asp Ser Asn Asp Trp Asn Val Arg Glu Ile Glu Lys Lys Ile
930 935 940

Glu Leu Ser Lys Lys Thr Glu Ile His Gly Pro Lys Gly Arg Glu Lys
945 950 955 960

Val Pro Lys Trp Ser Lys Glu Gln Phe Gln Ala Arg Gln His Lys Met
965 970 975

Ser Lys Pro Gln Arg Gln Asp Ser Arg Glu Ala Glu Lys Phe Lys Asp
980 985 990

Ile Asp Gln Thr Ile Arg Asn Leu Asp Lys Gln Leu Lys Glu Gly His
995 1000 1005

Asn Leu Asp Val Gly Glu Arg Gly Arg Asn Lys Val Ala Ser Ile
1010 1015 1020

Ala Gly Gln Phe Gly Lys Lys Asp Glu Ala Asn Ser Asp Glu Lys
1025 1030 1035

Asn Ala Gly Ser Ser Asn Ala Thr Thr Asn Thr Asn Asn Thr Val
1040 1045 1050

Ile Pro Lys Ser Ser Ser Lys Val Ala Leu Ala Phe Lys Lys Gln
1055 1060 1065

Ala Ala Ser Glu Lys Cys Arg Phe Cys Lys Gln Thr Val Tyr Leu
1070 1075 1080

Met Glu Lys Thr Thr Val Glu Gly Leu Val Leu His Arg Asn Cys
1085 1090 1095

Leu Lys Cys His His Cys His Thr Asn Leu Arg Leu Gly Gly Tyr
1100 1105 1110

Ala Phe Asp Arg Asp Asp Pro Gln Gly Arg Phe Tyr Cys Thr Gln
1115 1120 1125

His Phe Arg Leu Pro Pro Lys Pro Leu Pro Gln Arg Thr Asn Lys
1130 1135 1140

Ala Arg Lys Ser Ala Ala Ala Gln Pro Ala Ser Pro Ala Val Pro
1145 1150 1155

Pro Thr Ala Gly Ser Val Pro Thr Ala Ala Ala Thr Ser Glu His
1160 1165 1170

Met Asp Thr Thr Pro Pro Arg Asp Gln Val Asp Leu Leu Gln Thr
1175 1180 1185

Ser Arg Ala Asn Ala Ser Ala Asp Ala Met Ser Asp Asp Glu Ala
1190 1195 1200

Asn Val Ile Asp Glu His Glu Trp Ser Gly Arg Asn Phe Leu Pro
1205 1210 1215

Glu Ser Asn Asn Asp Ser Gln Ser Glu Leu Ser Ser Ser Asp Glu
1220 1225 1230

Ser Asp Thr Glu Ser Asp Ser Glu Met Phe Glu Glu Ala Asp Asp
1235 1240 1245

Ser Pro Phe Gly Ala Gln Thr Leu Gln Leu Ala Ser Asp Trp Ile
1250 1255 1260

Gly Lys Gln Tyr Cys Glu Asp Ser Asp Asp Ser Asp Asp Phe Tyr
1265 1270 1275

Asp Ser Ser Glu Gly Ile Ala Asp Asp Gly Lys Asp Asp Thr Glu
1280 1285 1290

Gly Glu Glu Phe Lys Lys Ala Arg Glu Leu Arg Arg Gln Glu Val
1295 1300 1305

Arg Leu Gln Pro Leu Pro Ala Asn Leu Pro Thr Asp Thr Glu Thr
1310 1315 1320

Glu Val Gln Thr Glu Ser Glu Ser Thr Ser Pro Asp Glu Val Glu
1325 1330 1335

Leu Asn Ser Ala Thr Glu Ile Ser Thr Asp Ser Glu Phe Asp Asn
1340 1345 1350

Asp Glu Ile Ile Arg Gln Ala Pro Lys Ile Phe Ile Asp Asp Thr
1355 1360 1365

His Leu Arg Lys Pro Thr Lys Val Gln Ile Lys Ser Thr Met Ile
1370 1375 1380

Gly Pro Asn Ala Ala Ser Ala Gly Leu His Gln Lys Gln Leu Ala
1385 1390 1395

Ala Arg Glu Lys Gly Gly Ser Tyr Leu Gln Lys Tyr Gln Pro Gln
1400 1405 1410

Pro Pro Leu Ser Gln Phe Lys Pro Leu Val Gln Val Asp Pro Thr
1415 1420 1425

Leu Leu Ile Gly Ser Gln Arg Ala Pro Leu Gln Asn Pro Arg Pro
1430 1435 1440

Gly Asp Tyr Leu Leu Asn Lys Thr Ala Ser Thr Glu Gly Ile Ala
1445 1450 1455

Ser Lys Lys Ser Leu Glu Leu Lys Lys Arg Tyr Leu Leu Gly Glu
1460 1465 1470

Pro Ala Asn Gly Asp Lys Ile Gln Lys Ser Gly Ser Thr Ser Val
1475 1480 1485

Leu Asp Ser Arg Ile Arg Ser Phe Gln Ser Asn Ile Ser Glu Cys
1490 1495 1500

Gln Lys Leu Leu Asn Pro Ser Ser Asp Ile Ser Ala Gly Met Arg
1505 1510 1515

Thr Phe Leu Asp Arg Thr Lys Leu Gly Glu Gly Ser Gln Thr Thr
1520 1525 1530

Pro Gly Gln Thr Asn Glu Leu Ile Arg Ser Ala Thr Ser Asn Val

1535 1540 1545

Ile Asn Asp Leu Arg Val Glu Leu Arg Ile Gln Lys Thr Asp Ser
1550 1555 1560

Ser His Ser Thr Asp Asn Glu Lys Glu Asn Val Phe Val Asn Cys
1565 1570 1575

Lys Asn Glu Leu Asn Lys Gly Met Glu Tyr Thr Asp Ala Val Asn
1580 1585 1590

Ala Thr Leu Leu Asp Gln Leu Ala Arg Lys Ser Ser Pro Thr Thr
1595 1600 1605

Pro Thr Asn Lys Thr Val Val Glu Val Ile Asp Leu Val Thr Pro
1610 1615 1620

Glu Lys Pro Ile Asp Ile Ile Asp Leu Thr Ala Leu Glu Thr Pro
1625 1630 1635

Lys Lys Gln Leu Val Asp Gly Ser Ala Met Asp Val Asp Glu Arg
1640 1645 1650

Leu Thr Pro Asp Ser Asn Lys Ile Ser Glu Leu Gln Gln Glu Val
1655 1660 1665

Lys Glu Glu Pro Lys Pro Asp Val Ser Arg Asp Val Lys Glu Cys
1670 1675 1680

Ile Pro Asp Ile Leu Gly His Ile Lys Glu Gly Thr Gly Ser Lys
1685 1690 1695

Glu Pro Gly Gly Glu Asp Gln Gln Ser Leu Leu Glu Gln Ser Asp
1700 1705 1710

Glu Glu Lys Arg Asp Ser Pro Glu Lys Asp Val Ala Glu His Glu
1715 1720 1725

Leu Tyr Glu Pro Asp Ser Val Gln Ile Gln Val Pro Asn Ile Pro
1730 1735 1740

Trp Glu Lys Ser Lys Pro Glu Val Met Ser Thr Thr Gly Ser Ser
1745 1750 1755

Gly Ser Ile Cys Ser Ser Ser Asp Ser Ser Ser Ile Glu Asp Ile
1760 1765 1770

Gln His Tyr Ile Leu Glu Ser Thr Thr Ser Pro Asp Thr Gln Thr
1775 1780 1785

Val Gly Gly Lys His Asn Val Pro Arg Leu Glu Val His Asp Thr
1790 1795 1800

Ser Gly Ala Leu Met Gln Val Asp Ser Leu Met Ile Val Asn Gly
1805 1810 1815

Lys Tyr Ile Gly Asp Pro Glu Asp Val Lys Phe Leu Asp Met Pro
1820 1825 1830

Ala Asn Val Ile Val Pro Pro Ala Pro Ala Leu Lys Thr Asn Glu
1835 1840 1845

Leu Asp Met Glu Asp Asp Gln Glu Ala Glu Ala Glu Pro Val Thr
1850 1855 1860

Ala Thr Pro Glu Pro Val Glu Cys Thr Val Ile Glu Ala Glu Arg
1865 1870 1875

Arg Val Thr Ala Pro Pro Pro Leu Pro Glu Met Gly Pro Pro Lys
1880 1885 1890

Leu Lys Phe Asp Ser Lys Asn Glu Asn Lys Ile Glu Ser Leu Lys
1895 1900 1905

Asn Leu Pro Leu Ile Val Glu Ser Asn Val Glu His Ser Gln Ala
1910 1915 1920

Val Lys Pro Ile Thr Leu Asn Leu Ser Asn Leu Ala Arg Thr Pro
1925 1930 1935

Asp Thr Pro Thr Thr Pro Thr Ala His Asp Ser Asp Lys Thr Pro
1940 1945 1950

Thr Gly Glu Ile Leu Ser Arg Gly Ser Asp Ser Glu Thr Glu His
1955 1960 1965

Thr Gly Thr Gly Gln Val Leu Thr Glu Thr Glu Leu Ser Asp Trp
1970 1975 1980

Thr Ala Asp Asp Cys Ile Ser Glu Asn Phe Val Asp Leu Glu Phe
1985 1990 1995

Ala Leu Asn Ser Asn Lys Gly Thr Ile Lys Arg Arg Lys Asp Arg
2000 2005 2010

Arg Arg Ser Gly Ala Ser Lys Leu Pro Ser Gly Asn Glu Val Ile
2015 2020 2025

His Glu Leu Ala Arg Gln Ala Pro Val Val Gln Met Asp Gly Ile
2030 2035 2040

Leu Ser Ala Ile Asp Ile Asp Asp Ile Glu Phe Met Asp Thr Gly
2045 2050 2055

Ser Glu Gly Ser Cys Ala Glu Ala Tyr Pro Ala Thr Asn Thr Ala
2060 2065 2070

Leu Ile Gln Asn Arg Gly Tyr Met Glu Tyr Ile Glu Ala Glu Pro
2075 2080 2085

Lys Lys Thr Thr Arg Lys Ala Ala Pro Pro Ser Ser Tyr Pro Gly
2090 2095 2100

Asn Leu Pro Pro Leu Met Thr Lys Arg Asp Glu Lys Leu Gly Val
2105 2110 2115

Asp Tyr Ile Glu Gln Gly Ala Tyr Ile Met His Asp Asp Ala Lys
2120 2125 2130

Thr Pro Val Asn Glu Val Ala Pro Ala Met Thr Gln Ser Leu Thr
2135 2140 2145

Asp Ser Ile Thr Leu Asn Glu Leu Asp Asp Asp Ser Met Ile Ile
2150 2155 2160

Ser Gln Thr Gln Pro Thr Thr Thr Glu Glu Ser Glu Ala Leu Thr
2165 2170 2175

Val Val Thr Ser Pro Leu Asp Thr Ser Ser Pro Arg Val Leu Asp
2180 2185 2190

Gln Phe Ala Ser Met Leu Ala Ala Gly Lys Gly Asp Ser Thr Pro
2195 2200 2205

Ser Ser Ser Glu Gln Gln Pro Lys Thr Ser Thr Val Thr Ser Ser
2210 2215 2220

Ser Thr Gly Pro Asn Ser Ser Thr Thr Gly Asn Val Ser Lys Glu
2225 2230 2235

Pro Gln Glu Glu Asp Leu Gln Ile Gln Phe Glu Tyr Val Arg Ala
2240 2245 2250

Leu Gln Gln Arg Ile Ser Gln Ile Ser Thr Gln Arg Arg Lys Ser
2255 2260 2265

Ser Lys Gly Glu Ala Pro Asn Leu Gln Leu Asn Ser Ser Ala Pro
2270 2275 2280

Val Ile Glu Ser Ala Glu Asp Pro Ala Lys Pro Ala Glu Glu Pro
2285 2290 2295

Leu Val Ser Met Arg Pro Arg Thr Thr Ser Ile Ser Gly Lys Val
2300 2305 2310

Pro Glu Ile Pro Thr Leu Ser Ser Lys Leu Glu Glu Ile Thr Lys
2315 2320 2325

Glu Arg Thr Lys Gln Lys Asp Leu Ile His Asp Leu Val Met Asp
2330 2335 2340

Lys Leu Gln Ser Lys Lys Gln Leu Asn Ala Glu Lys Arg Leu His
2345 2350 2355

Arg Ser Arg Gln Arg Ser Leu Leu Thr Ser Gly Tyr Ala Ser Gly
2360 2365 2370

Ser Ser Leu Ser Pro Thr Pro Lys Leu Ala Ala Ala Cys Ser Pro
2375 2380 2385

Gln Asp Ser Asn Cys Ser Ser Gln Ala His Tyr His Ala Ser Thr
2390 2395 2400

Ala Glu Glu Ala Pro Lys Pro Pro Ala Glu Arg Pro Leu Gln Lys
2405 2410 2415

Ser Ala Thr Ser Thr Tyr Val Ser Pro Tyr Arg Thr Val Gln Ala
2420 2425 2430

Pro Thr Arg Ser Ala Asp Leu Tyr Lys Pro Arg Pro Phe Ser Glu
2435 2440 2445

His Ile Asp Ser Asn Ala Leu Ala Gly Tyr Lys Leu Gly Lys Thr

2450 2455 2460

Ala Ser Phe Asn Gly Gly Lys Leu Gly Asp Phe Ala Lys Pro Ile
2465 2470 2475

Ala Pro Ala Arg Val Asn Arg Gly Gly Gly Val Ala Thr Ala Asp
2480 2485 2490

Ile Ala Asn Ile Ser Ala Ser Thr Glu Asn Leu Arg Ser Glu Ala
2495 2500 2505

Arg Ala Arg Ala Arg Leu Lys Ser Asn Thr Glu Leu Gly Leu Ser
2510 2515 2520

Pro Glu Glu Lys Met Gln Leu Ile Arg Ser Arg Leu His Tyr Asp
2525 2530 2535

Gln Asn Arg Ser Leu Lys Pro Lys Gln Leu Glu Glu Met Pro Ser
2540 2545 2550

Gly Asp Leu Ala Ala Arg Ala Arg Lys Met Ser Ala Ser Lys Ser
2555 2560 2565

Val Asn Asp Leu Ala Tyr Met Val Gly Gln Gln Gln Gln Gln Gln
2570 2575 2580

Val Glu Lys Asp Ala Val Leu Gln Ala Lys Ala Ala Asp Phe Thr
2585 2590 2595

Ser Asp Pro Asn Leu Ala Ser Gly Gly Gln Glu Lys Ala Gly Lys
2600 2605 2610

Thr Lys Ser Gly Arg Arg Pro Lys Asp Pro Glu Arg Arg Lys Ser
2615 2620 2625

Leu Ile Gln Ser Leu Ser Ser Phe Phe Gln Lys Gly Ser Gly Ser
2630 2635 2640

Ala Ala Ser Ser Ser Lys Glu Gln Gly Gly Ala Val Ala Ala Val
2645 2650 2655

His Ser Glu Gln Ser Glu Arg Pro Gly Thr Ser Ser Ser Gly Thr
2660 2665 2670

Pro Thr Ile Ser Asp Ala Ala Gly Gly Gly Gly Gly Gly Gly Gly
2675 2680 2685

Val Phe Ser Arg Phe Arg Ile Ser Pro Lys Ser Lys Glu Lys Ser
2690 2695 2700

Lys Ser Cys Phe Asp Leu Arg Asn Phe Gly Phe Gly Asp Lys Asp
2705 2710 2715

Met Leu Val Cys Asn Ala Ala Ser Pro Ala Gly Ala Thr Ser Ala
2720 2725 2730

Ser Gln Lys Asn His Ser Gln Glu Tyr Leu Asn Thr Thr Asn Asn
2735 2740 2745

Ser Arg Tyr Arg Lys Gln Thr Asn Thr Ala Lys Pro Lys Pro Glu
2750 2755 2760

Ser Phe Ser Ser Ser Ser Pro Gln Leu Tyr Ile His Lys Pro His
2765 2770 2775

His Leu Ala Ala Ala His Pro Ser Ala Leu Asp Asp Gln Thr Pro
2780 2785 2790

Pro Pro Ile Pro Pro Leu Pro Leu Asn Tyr Gln Arg Ser Asp Asp
2795 2800 2805

Glu Ser Tyr Ala Asn Glu Thr Arg Glu His Lys Lys Gln Arg Ala
2810 2815 2820

Ile Ser Lys Ala Ser Arg Gln Ala Glu Leu Lys Arg Leu Arg Ile
2825 2830 2835

Ala Gln Glu Ile Gln Arg Glu Gln Glu Glu Ile Glu Val Gln Leu
2840 2845 2850

Lys Asp Leu Glu Ala Arg Gly Val Leu Ile Glu Lys Ala Leu Arg
2855 2860 2865

Gly Glu Ala Gln Asn Ile Glu Asn Leu Asp Ala Thr Lys Asp Asn
2870 2875 2880

Asp Glu Lys Leu Leu Lys Glu Leu Leu Glu Ile Trp Arg Asn Ile
2885 2890 2895

Thr Ala Leu Lys Lys Arg Asp Glu Glu Leu Thr Ile Arg Gln Gln
2900 2905 2910

Glu Leu Gln Leu Glu Tyr Arg His Ala Gln Leu Lys Glu Glu Leu
2915 2920 2925

Asn Leu Arg Leu Ser Cys Asn Lys Leu Asp Lys Ser Ser Ala Asp
2930 2935 2940

Val Ala Ala Glu Gly Ala Ile Leu Asn Glu Met Leu Glu Ile Val
2945 2950 2955

Ala Lys Arg Ala Ala Leu Arg Pro Thr Ala Ser Gln Leu Asp Leu
2960 2965 2970

Thr Ala Ala Gly Ser Ala Ser Thr Ser Ala Glu Ala Thr Gly Ile
2975 2980 2985

Lys Leu Thr Gly Gln Pro His Asp His Glu Glu Ser Ile Ile
2990 2995 3000

<210> 11
<211> 8205
<212> DNA
<213> Drosophila

<400> 11
atgagccgcc aacaccagcg gcaccaccag cagcatcacc acctgccgcc gcaccagcaa 60

ccgcagcagc agatgccgca acaacagcag cagctgacgg cgcagcagca gcaacaacag 120

cagctgctga tggcggagca cgcggcggcc gcggaggcgg cggagctatt cgacctgctg 180

tgcgtggcca caacgatgcg ccagatcctg gcgctccatc gggccatgtg cgaggctgtg 240

ggattgagac cctcgcctct gaacgacttc tacccacggc taaaggccaa ggtgcgttcg 300

tggaaggcgc aggccctgtg gaagaagttc gacgccagag ctgcccatag agtctacggc 360

aagggagctg cctgtactgg cacacgcgtc ctggtcatcg gagcagggcc ctgtggactg 420

cgcaccgcca tcgaggccca actgctgggc gccaaggtgg tggtgctgga gaaacgcgat 480

cgcatcaccc ggaacaatgt gctccatctg tggccattcg tcatcacgga tctgcgcaac 540

ttgggcgcaa agaagttcta cggcaagttt tgcgccggct ccatcgatca catctccatt 600

cggcagctgc agtgcatgct gctcaaggtg gcgctgctcc tgggcgtaga gatccacgag 660

ggagtcagtt ttgatcacgc tgtagagccc tctggcgatg gcggcggatg gagggcagct 720

gttactcccg cagatcatcc tgtatctcac tacgaattcg atgtgttgat cggagcggat 780

ggcaagcgga atatgctgga ctttaggagg aaggagttcc gcgggaagct ggccatcgct 840

attacagcga actttatcaa caagaagacg gaggcggagg ctaaagtaga ggagatcagt 900

gggggtggctt tcatcttcaa ccaggccttc ttcaaggagc tgtacgggaa gacgggcatc 960
gacctggaaa acatcgtcta ctacaaggac gagacgcact acttcgtgat gacggccaag 1020
aagcacagtc taattgacaa gggcgttatt atcgaggata tggccgatcc cggcgagctt 1080
ctcgccccag ccaatgtgga tacacaaaag ctgcacgact atgcacgcga ggctgcggag 1140
ttctccaccc aataccaaat gccaaacctg gagttcgctg ttaatcacta cggcaaacca 1200
gatgtggcca tgttcgactt cacatcgatg tttgccgccg agatgtcctg tcgggtgatt 1260
gtgcgcaaag gagctcgcct gatgcagtgc ctcgtgggtg acagtctgct cgagccgttt 1320
tggcccactg gatcgggttg tgcccgtgga ttcttatcca gcatggatgc tgcctatgcc 1380
atcaagcttt ggtccaaccc gcagaacagc acacttggcg ttctggcgca gcgcgaaagc 1440
atctaccggc tgcttaacca gaccacgccg gacaccctgc agcgggacat cagtgcctat 1500
accgtggatc cggccacgcg ctatccgaat ctgaacaggg agtcggtcaa tagctggcag 1560
gtcaaacatc tggtcgacac ggacgacccg tccattctgg agcagacctt catggacacg 1620
catgctctgc agaccccgca tttggacaca ccgggcagac gcaagcgacg cagtggagac 1680
ttgctgcccc agggtgccac gttgctgaga tggataagtg cccagctgca ttcctatcag 1740
tttattcccg aactcaagga ggcttcggat gtgttccgga atggacgcgt tctgtgtgcg 1800
cttatcaatc gctatcgtcc tgatctcatc gactacgctg ccaccaagga catgagtccc 1860
gtggagtgca atgagctgtc attcgccgtc ctagagcgcg aactccacat cgatcgcgtc 1920
atgagtgcca aacagtcgct ggacttgacc gagctggagt cgcgaatctg gctcaactat 1980
ttggaccaga tctgcgactt gtttcgcggc gagatccccc atatcaagca ccccaagatg 2040
gactttagcg atttgcgcca gaagtatcgt atcaaccata cgcatgccca acccgacttc 2100
tccaagctgc tggcaacgaa acccaaggcc aagtcgccga tgcaggatgc tgtggacata 2160
cccacgacag tgcagcggcg ctcggtgctc gaggaggagc gagccaagcg gcagcgtcgc 2220
cacgagcagc ttcttaacat cggtggaggg gcagcaggag ccgccgccgg agttgccgga 2280
agcgggacag gaaccacaac gcagggtcaa aacgatacgc cacgccggtc caagaagcgc 2340
cgtcaggttg acaaaaccgc caatattagt tccaaggtgg cgctggcctt taaaaagcag 2400
gctgcctccg aaaagtgccg cttctgtaag caaaccgtct acctgatgga gaagaccacc 2460
gtggagggat tggttctgca tcgcaattgc cttaagtgcc accactgcca caccaacttg 2520
cgtctgggag gctacgcctt tgatcgggac gatccgcagg gccgatttta ctgcacccaa 2580
cacttccggt tgccacccaa accgctgccg cagcgcacca acaaagccag gaaatccgct 2640
gccgctcaac ccgcctcgcc tgctgtacca ccaactgcgg gatccgtacc cactgcagct 2700
gccacatcgg agcatatgga caccactcca cccagggacc aggtggacct actggagacc 2760

tcgcgagcaa atgcctctgc cgatgccatg tccgatgatg aagccaatgt tatcgatgag 2820
cacgaatggt ctggtcgcaa cttcttgccc gagtccaaca acgattccca atcagagcta 2880
tccagttcag atgagtcgga tacggaatcg gattcggaga tgtttgagga ggcggatgat 2940
tcgccgtttg gtgctcagac cctccagctg gcgtcggatt ggattggaaa gcaatactgt 3000
gaggacagtg atgattctga cgatttctac gactcaagtg aaggtattgc ggatgacggc 3060
aaagatgaca ccgagggtga ggaattcaag aaggcccgcg aattgaggcg ccaggaagtt 3120
cgcctgcagc cgttgcccgc caatctgccc acagatacgg agaccgaggt tcaaaccgag 3180
tccgagagca cttcacccga cgaagtggag ctcaattctg ccactgagat atccaccgac 3240
tctgagtttg acaacgatga gattatacgc caggcgccca aaatcttcat cgatgacacc 3300
catctaagga agcccaccaa ggttcagatc aagtccacca tgatcggacc caatgcagct 3360
tccgccggac tccatcagaa gcagttggcg gcgcgtgaga agggcggcag ctacctccag 3420
aagtaccaac cacaaccgcc actgtcacag tttaaaccgt tggtccaggt ggatcccacc 3480
ctgctcattg gcagccagcg cgctcctctt cagaatccac ggccaggaga ctacttgcta 3540
aacaagacgg ccagtacgga gggtatcgcc tcaaaaaaga gcctggagct aaaaaagcgc 3600
tatctgctgg gtgagccggc caatggcgat aagatccaga agtccggatc cacttcagtg 3660
ctggattcac gcattcgcag cttccagtcg aacatatcgg agtgccagaa gcttttgaat 3720
cccagcagcg acataagtgc cggcatgcga accttcctcg atcgcacaaa gttgggcgaa 3780
ggcagccaga cgacacccgg acagacgaac gaactaatcc gttccgccac cagcaatgtg 3840
attaacgatc tgcgcgtgga gcttcggata cagaaaactg actccagcca ctccacggac 3900
aacgagaagg aaaacgtttt cgtgaactgt aagaacgagc tgaacaaggg gatggaatac 3960
acggatgcgg tcaatgccac gctgctggac cagctggcca gaaaaagttc acccaccacg 4020
ccgacgaata agacggtggt cgaggttatt gacctggtta cacctgagaa gccaattgac 4080
attatcgatc taacggcact ggaaacgccg aaaaagcagt tggtcgatgg tagcgccatg 4140
gatgtagatg aacgcctcac acccgatagc aacaaaatca gcgaactgca gcaggaagtg 4200
aaggaggaac ccaagccgga tgtctctagg gatgtgaaag aatgcatacc agatatactg 4260
ggacacatta aggagggaac gggatcgaag gagccaggtg gagaggacca acagagcctg 4320
ctggagcagt cggacgaaga gaagcgcgac tcaccggaaa aggatgtggc cgaacatgag 4380
ctttatgaac cggacagtgt gcagatccag gtgcccaata tcccatggga aaaaagcaag 4440
ccggaggtca tgtctaccac cggcagcagt ggctccatct gctcaagctc agactcttct 4500
agtattgaag acatccagca ctacattttg gagtccacaa ctagtccaga tactcagaca 4560

gttggcggaa agcacaatgt gccccgtttg gaggtgcacg acacaagtgg tgccctgatg 4620
caggtggaca gcctgatgat tgtgaacgga aagtatattg gggatcccga ggatgtcaag 4680
ttcttggata tgccggccaa tgttattgtt ccgccagcac cggcgcttaa aacgaatgag 4740
ctggatatgg aggatgacca agaggcggag gcggaaccag taactgctac tccggagccg 4800
gtggaatgta cggtcatcga ggctgagcgc cgtgttactg ctccccctcc tttgccagag 4860
atgggtccac ccaaactgaa gttcgatagc aaaaatgaga acaagatcga gagcttgaag 4920
aatcttccgt tgatcgtaga gagcaatgtg gagcacagtc aggcagtgaa acccattact 4980
cttaacttaa gcaatctggc caggacgccg gatacaccaa ccacgcccac ggcgcacgat 5040
agcgataaaa cacccactgg ggaaattctc tcgcgaggat ctgactcaga aaccgagcac 5100
actggcactg gtcaggtact aacggagacg gaactctccg actggacggc cgacgactgt 5160
atctcggaga actttgttga cttggagttc gcgcttaact ctaacaaggg tacgataaaa 5220
cggcgcaagg atcgacgacg cagtggagca agcaaacttc ccagtggcaa cgaggtaatc 5280
cacgagctgg ccaggcaggc gccagtggtg caaatggatg gaattcttag tgccatcgac 5340
attgatgaca tagagttcat ggacacgggt tcggagggtt cttgtgctga agcttatccc 5400
gcaacaaata cagctctcat tcagaataga ggttacatgg agtacatcga ggcggagccg 5460
aaaaagacga cccgcaaggc agctccacca tccagttacc caggaaattt accgccttta 5520
atgacgaagc gggacgagaa actgggcgtt gattacattg agcagggggc gtacataatg 5580
cacgatgatg caaagacgcc tgtgaatgag gtggctcctg ccatgaccca gtcgctaact 5640
gactcaatca cgctcaatga actggatgat gacagtatga taatatccca aacccagcca 5700
acgacaacgg aggaaagtga ggcactgacg gtggtcacca gtccacttga cacgtcctcg 5760
cccagggttc tcgatcaatt tgcatccatg ttggcggcgg gaaaaggtga ctccacaccc 5820
agtagctcag agcaacaacc aaagacgtct acggtgacga gcagcagcac tgggcccaac 5880
tcctcgacaa caggaaacgt ctcgaaggag ccgcaggagg aggacctgca aatccagttt 5940
gagtatgttc gagcactgca gcagcggata tcgcagatca gcacccaacg gcgtaagagc 6000
tctaagggag aggcacctaa cctgcagcta aacagtagcg cacctgtgat agaatcagcc 6060
gaggatccgg ccaagcccgc agaggagcct ctggtctcaa tgcgaccgcg gaccaccagc 6120
atttccggaa aggtaccgga gatacccaca cttagcagca agctggaaga gataaccaaa 6180
gaacgcacta agcaaaagga tctgattcac gacctagtca tggacaagtt gcagtcgaag 6240
aagcagctaa acgctgagaa gcgtctgcac cggagtcgac agcgcagttt gctgaccagt 6300
ggctatgcca gtggctccag ccttagtccg acgcccaagc tggctgctgc ttgcagtccg 6360
caggattcca actgctctag ccaagcgcac taccacgcct ccacggcgga ggaggccccg 6420

aagccgccgg cggaaaggcc gttgcagaag tccgccacgt ccacctatgt gtcgccttat 6480
cgcactgtcc aagcgcccac acgtagtgct gatctctata agccgcgccc cttcagcgaa 6540
cacatcgatt cgaacgctct ggcgggttac aagctcggca agacggcctc gtttaatggc 6600
ggcaagttgg gcgactttgc gaaacccatt gccccggcga gagttaaccg aggaggaggt 6660
gtcgcgaccg cggatatagc caatatttcc gcgtcgacgg agaacctaag aagcgaggcc 6720
agggccaggg ctcgtcttaa gtctaacaca gagctgggcc ttagtcccga ggaaaagatg 6780
cagctaatac gttcaagatt gcactacgac caaaacagat ctctgaagcc gaagcaactg 6840
gaggagatgc catccgggga tctggcggca cgtgcccgca aaatgagtgc ctcgaagagc 6900
gtcaatgatc tggcctacat ggtgggacag cagcagcagc agcaggttga gaaggatgcc 6960
gtgctccaag ccaaggcggc tgactttaca tccgatccca atttggcgtc cggtggtcag 7020
gagaaggcag gcaaaactaa gtccggacgc aggccaaagg atccggagcg gcgtaagagt 7080
ctcatacagt cgttgtccag cttcttccaa aagggatctg gatccgcggc ctccagttcc 7140
aaggagcagg gcggcgctgt ggctgccgtc cactctgaac agtcagagcg accaggcacc 7200
agcagcagcg gcacgcccac aatatcggat gcggcgggtg gaggcggagg aggaggtggc 7260
gtcttcagca gattccgcat ctcgcccaag tccaaggaga agtcaaagtc ttgctttgat 7320
ctgaggaatt tcggttttgg tgacaaggat atgctggtct gcaatgcagc atctccagca 7380
ggagccacat ccgcatcaca gaaaaatcac tcgcaagagt atctgaacac cacgaacaac 7440
agtcgctatc gaaagcaaac gaacactgcg aaaccgaaac ccgaatcgtt ctcttcatcc 7500
agtccgcagc tctatataca caagccccac cacctggccg cagctcatcc cagtgccctg 7560
gacgaccaga caccaccacc cataccgcct cttccactga attatcagag atccgatgat 7620
gagagctacg ctaacgagac acgagagcat aagaagcaac gtgccatatc gaaggcttca 7680
cgacaagctg agctcaagcg attgcgaatc gctcaagaga ttcagcggga acaggaggag 7740
atcgaggtgc aactgaagga tctggaggca cgcggcgtgc ttattgagaa ggccttgcga 7800
ggcgaggcgc agaatattga aaacctggat gcgacaaagg acaacgacga gaagctactt 7860
aaggaacttt tggagatttg gcgcaacatc acagcactca agaaacgcga tgaggaactg 7920
actataaggc aacaggaact gcaactggag tatcggcatg cccagctgaa ggaagagctc 7980
aatctgcgct tgtcctgcaa caaactggac aaaagctctg ccgatgtggc cgccgaggga 8040
gcaattctca acgagatgct ggaaattgtc gccaagcgag ccgcccctacg acccacagcc 8100
tcccagctcg acctcacggc agcgggatca gcatccacgt ccgccgaggc aacgggcatt 8160
aagctgacgg gacaaccgca tgaccacgaa gaatcgatca tttga 8205

<210> 12
<211> 2734
<212> PRT
<213> Drosophila

<400> 12

Met Ser Arg Gln His Gln Arg His His Gln Gln His His His Leu Pro
1 5 10 15

Pro His Gln Gln Pro Gln Gln Gln Met Pro Gln Gln Gln Gln Gln Leu
20 25 30

Thr Ala Gln Gln Gln Gln Gln Gln Gln Leu Leu Met Ala Glu His Ala
35 40 45

Ala Ala Ala Glu Ala Ala Glu Leu Phe Asp Leu Leu Cys Val Ala Thr
50 55 60

Thr Met Arg Gln Ile Leu Ala Leu His Arg Ala Met Cys Glu Ala Val
65 70 75 80

Gly Leu Arg Pro Ser Pro Leu Asn Asp Phe Tyr Pro Arg Leu Lys Ala
85 90 95

Lys Val Arg Ser Trp Lys Ala Gln Ala Leu Trp Lys Lys Phe Asp Ala
100 105 110

Arg Ala Ala His Arg Val Tyr Gly Lys Gly Ala Ala Cys Thr Gly Thr
115 120 125

Arg Val Leu Val Ile Gly Ala Gly Pro Cys Gly Leu Arg Thr Ala Ile
130 135 140

Glu Ala Gln Leu Leu Gly Ala Lys Val Val Val Leu Glu Lys Arg Asp
145 150 155 160

Arg Ile Thr Arg Asn Asn Val Leu His Leu Trp Pro Phe Val Ile Thr
165 170 175

Asp Leu Arg Asn Leu Gly Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala
180 185 190

Gly Ser Ile Asp His Ile Ser Ile Arg Gln Leu Gln Cys Met Leu Leu
195 200 205

Lys Val Ala Leu Leu Leu Gly Val Glu Ile His Glu Gly Val Ser Phe

210 215 220

Asp His Ala Val Glu Pro Ser Gly Asp Gly Gly Gly Trp Arg Ala Ala
225 230 235 240

Val Thr Pro Ala Asp His Pro Val Ser His Tyr Glu Phe Asp Val Leu
245 250 255

Ile Gly Ala Asp Gly Lys Arg Asn Met Leu Asp Phe Arg Arg Lys Glu
260 265 270

Phe Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Lys
275 280 285

Lys Thr Glu Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe
290 295 300

Ile Phe Asn Gln Ala Phe Phe Lys Glu Leu Tyr Gly Lys Thr Gly Ile
305 310 315 320

Asp Leu Glu Asn Ile Val Tyr Tyr Lys Asp Glu Thr His Tyr Phe Val
325 330 335

Met Thr Ala Lys Lys His Ser Leu Ile Asp Lys Gly Val Ile Ile Glu
340 345 350

Asp Met Ala Asp Pro Gly Glu Leu Leu Ala Pro Ala Asn Val Asp Thr
355 360 365

Gln Lys Leu His Asp Tyr Ala Arg Glu Ala Ala Glu Phe Ser Thr Gln
370 375 380

Tyr Gln Met Pro Asn Leu Glu Phe Ala Val Asn His Tyr Gly Lys Pro
385 390 395 400

Asp Val Ala Met Phe Asp Phe Thr Ser Met Phe Ala Ala Glu Met Ser
405 410 415

Cys Arg Val Ile Val Arg Lys Gly Ala Arg Leu Met Gln Cys Leu Val
420 425 430

Gly Asp Ser Leu Leu Glu Pro Phe Trp Pro Thr Gly Ser Gly Cys Ala
435 440 445

Arg Gly Phe Leu Ser Ser Met Asp Ala Ala Tyr Ala Ile Lys Leu Trp
450 455 460

```
Ser Asn Pro Gln Asn Ser Thr Leu Gly Val Leu Ala Gln Arg Glu Ser
465                 470                 475                 480

Ile Tyr Arg Leu Leu Asn Gln Thr Thr Pro Asp Thr Leu Gln Arg Asp
                485                 490                 495

Ile Ser Ala Tyr Thr Val Asp Pro Ala Thr Arg Tyr Pro Asn Leu Asn
            500                 505                 510

Arg Glu Ser Val Asn Ser Trp Gln Val Lys His Leu Val Asp Thr Asp
        515                 520                 525

Asp Pro Ser Ile Leu Glu Gln Thr Phe Met Asp Thr His Ala Leu Gln
    530                 535                 540

Thr Pro His Leu Asp Thr Pro Gly Arg Arg Lys Arg Arg Ser Gly Asp
545                 550                 555                 560

Leu Leu Pro Gln Gly Ala Thr Leu Leu Arg Trp Ile Ser Ala Gln Leu
                565                 570                 575

His Ser Tyr Gln Phe Ile Pro Glu Leu Lys Glu Ala Ser Asp Val Phe
            580                 585                 590

Arg Asn Gly Arg Val Leu Cys Ala Leu Ile Asn Arg Tyr Arg Pro Asp
        595                 600                 605

Leu Ile Asp Tyr Ala Ala Thr Lys Asp Met Ser Pro Val Glu Cys Asn
    610                 615                 620

Glu Leu Ser Phe Ala Val Leu Glu Arg Glu Leu His Ile Asp Arg Val
625                 630                 635                 640

Met Ser Ala Lys Gln Ser Leu Asp Leu Thr Glu Leu Glu Ser Arg Ile
                645                 650                 655

Trp Leu Asn Tyr Leu Asp Gln Ile Cys Asp Leu Phe Arg Gly Glu Ile
            660                 665                 670

Pro His Ile Lys His Pro Lys Met Asp Phe Ser Asp Leu Arg Gln Lys
        675                 680                 685

Tyr Arg Ile Asn His Thr His Ala Gln Pro Asp Phe Ser Lys Leu Leu
    690                 695                 700
```

Ala Thr Lys Pro Lys Ala Lys Ser Pro Met Gln Asp Ala Val Asp Ile
705 710 715 720

Pro Thr Thr Val Gln Arg Arg Ser Val Leu Glu Glu Glu Arg Ala Lys
725 730 735

Arg Gln Arg Arg His Glu Gln Leu Leu Asn Ile Gly Gly Gly Ala Ala
740 745 750

Gly Ala Ala Ala Gly Val Ala Gly Ser Gly Thr Gly Thr Thr Thr Gln
755 760 765

Gly Gln Asn Asp Thr Pro Arg Arg Ser Lys Lys Arg Arg Gln Val Asp
770 775 780

Lys Thr Ala Asn Ile Ser Ser Lys Val Ala Leu Ala Phe Lys Lys Gln
785 790 795 800

Ala Ala Ser Glu Lys Cys Arg Phe Cys Lys Gln Thr Val Tyr Leu Met
805 810 815

Glu Lys Thr Thr Val Glu Gly Leu Val Leu His Arg Asn Cys Leu Lys
820 825 830

Cys His His Cys His Thr Asn Leu Arg Leu Gly Gly Tyr Ala Phe Asp
835 840 845

Arg Asp Asp Pro Gln Gly Arg Phe Tyr Cys Thr Gln His Phe Arg Leu
850 855 860

Pro Pro Lys Pro Leu Pro Gln Arg Thr Asn Lys Ala Arg Lys Ser Ala
865 870 875 880

Ala Ala Gln Pro Ala Ser Pro Ala Val Pro Pro Thr Ala Gly Ser Val
885 890 895

Pro Thr Ala Ala Ala Thr Ser Glu His Met Asp Thr Thr Pro Pro Arg
900 905 910

Asp Gln Val Asp Leu Leu Gln Thr Ser Arg Ala Asn Ala Ser Ala Asp
915 920 925

Ala Met Ser Asp Asp Glu Ala Asn Val Ile Asp Glu His Glu Trp Ser
930 935 940

Gly Arg Asn Phe Leu Pro Glu Ser Asn Asn Asp Ser Gln Ser Glu Leu
945 950 955 960

Ser Ser Ser Asp Glu Ser Asp Thr Glu Ser Asp Ser Glu Met Phe Glu
965 970 975

Glu Ala Asp Asp Ser Pro Phe Gly Ala Gln Thr Leu Gln Leu Ala Ser
980 985 990

Asp Trp Ile Gly Lys Gln Tyr Cys Glu Asp Ser Asp Asp Ser Asp Asp
995 1000 1005

Phe Tyr Asp Ser Ser Glu Gly Ile Ala Asp Asp Gly Lys Asp Asp
1010 1015 1020

Thr Glu Gly Glu Glu Phe Lys Lys Ala Arg Glu Leu Arg Arg Gln
1025 1030 1035

Glu Val Arg Leu Gln Pro Leu Pro Ala Asn Leu Pro Thr Asp Thr
1040 1045 1050

Glu Thr Glu Val Gln Thr Glu Ser Glu Ser Thr Ser Pro Asp Glu
1055 1060 1065

Val Glu Leu Asn Ser Ala Thr Glu Ile Ser Thr Asp Ser Glu Phe
1070 1075 1080

Asp Asn Asp Glu Ile Ile Arg Gln Ala Pro Lys Ile Phe Ile Asp
1085 1090 1095

Asp Thr His Leu Arg Lys Pro Thr Lys Val Gln Ile Lys Ser Thr
1100 1105 1110

Met Ile Gly Pro Asn Ala Ala Ser Ala Gly Leu His Gln Lys Gln
1115 1120 1125

Leu Ala Ala Arg Glu Lys Gly Gly Ser Tyr Leu Gln Lys Tyr Gln
1130 1135 1140

Pro Gln Pro Pro Leu Ser Gln Phe Lys Pro Leu Val Gln Val Asp
1145 1150 1155

Pro Thr Leu Leu Ile Gly Ser Gln Arg Ala Pro Leu Gln Asn Pro
1160 1165 1170

Arg Pro Gly Asp Tyr Leu Leu Asn Lys Thr Ala Ser Thr Glu Gly

1175 1180 1185

Ile Ala Ser Lys Lys Ser Leu Glu Leu Lys Lys Arg Tyr Leu Leu
1190 1195 1200

Gly Glu Pro Ala Asn Gly Asp Lys Ile Gln Lys Ser Gly Ser Thr
1205 1210 1215

Ser Val Leu Asp Ser Arg Ile Arg Ser Phe Gln Ser Asn Ile Ser
1220 1225 1230

Glu Cys Gln Lys Leu Leu Asn Pro Ser Ser Asp Ile Ser Ala Gly
1235 1240 1245

Met Arg Thr Phe Leu Asp Arg Thr Lys Leu Gly Glu Gly Ser Gln
1250 1255 1260

Thr Thr Pro Gly Gln Thr Asn Glu Leu Ile Arg Ser Ala Thr Ser
1265 1270 1275

Asn Val Ile Asn Asp Leu Arg Val Glu Leu Arg Ile Gln Lys Thr
1280 1285 1290

Asp Ser Ser His Ser Thr Asp Asn Glu Lys Glu Asn Val Phe Val
1295 1300 1305

Asn Cys Lys Asn Glu Leu Asn Lys Gly Met Glu Tyr Thr Asp Ala
1310 1315 1320

Val Asn Ala Thr Leu Leu Asp Gln Leu Ala Arg Lys Ser Ser Pro
1325 1330 1335

Thr Thr Pro Thr Asn Lys Thr Val Val Glu Val Ile Asp Leu Val
1340 1345 1350

Thr Pro Glu Lys Pro Ile Asp Ile Ile Asp Leu Thr Ala Leu Glu
1355 1360 1365

Thr Pro Lys Lys Gln Leu Val Asp Gly Ser Ala Met Asp Val Asp
1370 1375 1380

Glu Arg Leu Thr Pro Asp Ser Asn Lys Ile Ser Glu Leu Gln Gln
1385 1390 1395

Glu Val Lys Glu Glu Pro Lys Pro Asp Val Ser Arg Asp Val Lys
1400 1405 1410

Glu Cys Ile Pro Asp Ile Leu Gly His Ile Lys Glu Gly Thr Gly
1415 1420 1425

Ser Lys Glu Pro Gly Gly Glu Asp Gln Gln Ser Leu Leu Glu Gln
1430 1435 1440

Ser Asp Glu Glu Lys Arg Asp Ser Pro Glu Lys Asp Val Ala Glu
1445 1450 1455

His Glu Leu Tyr Glu Pro Asp Ser Val Gln Ile Gln Val Pro Asn
1460 1465 1470

Ile Pro Trp Glu Lys Ser Lys Pro Glu Val Met Ser Thr Thr Gly
1475 1480 1485

Ser Ser Gly Ser Ile Cys Ser Ser Ser Asp Ser Ser Ser Ile Glu
1490 1495 1500

Asp Ile Gln His Tyr Ile Leu Glu Ser Thr Thr Ser Pro Asp Thr
1505 1510 1515

Gln Thr Val Gly Gly Lys His Asn Val Pro Arg Leu Glu Val His
1520 1525 1530

Asp Thr Ser Gly Ala Leu Met Gln Val Asp Ser Leu Met Ile Val
1535 1540 1545

Asn Gly Lys Tyr Ile Gly Asp Pro Glu Asp Val Lys Phe Leu Asp
1550 1555 1560

Met Pro Ala Asn Val Ile Val Pro Pro Ala Pro Ala Leu Lys Thr
1565 1570 1575

Asn Glu Leu Asp Met Glu Asp Asp Gln Glu Ala Glu Ala Glu Pro
1580 1585 1590

Val Thr Ala Thr Pro Glu Pro Val Glu Cys Thr Val Ile Glu Ala
1595 1600 1605

Glu Arg Arg Val Thr Ala Pro Pro Pro Leu Pro Glu Met Gly Pro
1610 1615 1620

Pro Lys Leu Lys Phe Asp Ser Lys Asn Glu Asn Lys Ile Glu Ser
1625 1630 1635

Leu Lys Asn Leu Pro Leu Ile Val Glu Ser Asn Val Glu His Ser
1640 1645 1650

Gln Ala Val Lys Pro Ile Thr Leu Asn Leu Ser Asn Leu Ala Arg
1655 1660 1665

Thr Pro Asp Thr Pro Thr Thr Pro Thr Ala His Asp Ser Asp Lys
1670 1675 1680

Thr Pro Thr Gly Glu Ile Leu Ser Arg Gly Ser Asp Ser Glu Thr
1685 1690 1695

Glu His Thr Gly Thr Gly Gln Val Leu Thr Glu Thr Glu Leu Ser
1700 1705 1710

Asp Trp Thr Ala Asp Asp Cys Ile Ser Glu Asn Phe Val Asp Leu
1715 1720 1725

Glu Phe Ala Leu Asn Ser Asn Lys Gly Thr Ile Lys Arg Arg Lys
1730 1735 1740

Asp Arg Arg Arg Ser Gly Ala Ser Lys Leu Pro Ser Gly Asn Glu
1745 1750 1755

Val Ile His Glu Leu Ala Arg Gln Ala Pro Val Val Gln Met Asp
1760 1765 1770

Gly Ile Leu Ser Ala Ile Asp Ile Asp Asp Ile Glu Phe Met Asp
1775 1780 1785

Thr Gly Ser Glu Gly Ser Cys Ala Glu Ala Tyr Pro Ala Thr Asn
1790 1795 1800

Thr Ala Leu Ile Gln Asn Arg Gly Tyr Met Glu Tyr Ile Glu Ala
1805 1810 1815

Glu Pro Lys Lys Thr Thr Arg Lys Ala Ala Pro Pro Ser Ser Tyr
1820 1825 1830

Pro Gly Asn Leu Pro Pro Leu Met Thr Lys Arg Asp Glu Lys Leu
1835 1840 1845

Gly Val Asp Tyr Ile Glu Gln Gly Ala Tyr Ile Met His Asp Asp
1850 1855 1860

Ala Lys Thr Pro Val Asn Glu Val Ala Pro Ala Met Thr Gln Ser
1865 1870 1875

Leu Thr Asp Ser Ile Thr Leu Asn Glu Leu Asp Asp Asp Ser Met
1880 1885 1890

Ile Ile Ser Gln Thr Gln Pro Thr Thr Thr Glu Glu Ser Glu Ala
1895 1900 1905

Leu Thr Val Val Thr Ser Pro Leu Asp Thr Ser Ser Pro Arg Val
1910 1915 1920

Leu Asp Gln Phe Ala Ser Met Leu Ala Ala Gly Lys Gly Asp Ser
1925 1930 1935

Thr Pro Ser Ser Ser Glu Gln Gln Pro Lys Thr Ser Thr Val Thr
1940 1945 1950

Ser Ser Ser Thr Gly Pro Asn Ser Ser Thr Thr Gly Asn Val Ser
1955 1960 1965

Lys Glu Pro Gln Glu Glu Asp Leu Gln Ile Gln Phe Glu Tyr Val
1970 1975 1980

Arg Ala Leu Gln Gln Arg Ile Ser Gln Ile Ser Thr Gln Arg Arg
1985 1990 1995

Lys Ser Ser Lys Gly Glu Ala Pro Asn Leu Gln Leu Asn Ser Ser
2000 2005 2010

Ala Pro Val Ile Glu Ser Ala Glu Asp Pro Ala Lys Pro Ala Glu
2015 2020 2025

Glu Pro Leu Val Ser Met Arg Pro Arg Thr Thr Ser Ile Ser Gly
2030 2035 2040

Lys Val Pro Glu Ile Pro Thr Leu Ser Ser Lys Leu Glu Glu Ile
2045 2050 2055

Thr Lys Glu Arg Thr Lys Gln Lys Asp Leu Ile His Asp Leu Val
2060 2065 2070

Met Asp Lys Leu Gln Ser Lys Lys Gln Leu Asn Ala Glu Lys Arg
2075 2080 2085

Leu His Arg Ser Arg Gln Arg Ser Leu Leu Thr Ser Gly Tyr Ala

2090 2095 2100

Ser Gly Ser Ser Leu Ser Pro Thr Pro Lys Leu Ala Ala Ala Cys
2105 2110 2115

Ser Pro Gln Asp Ser Asn Cys Ser Ser Gln Ala His Tyr His Ala
2120 2125 2130

Ser Thr Ala Glu Glu Ala Pro Lys Pro Pro Ala Glu Arg Pro Leu
2135 2140 2145

Gln Lys Ser Ala Thr Ser Thr Tyr Val Ser Pro Tyr Arg Thr Val
2150 2155 2160

Gln Ala Pro Thr Arg Ser Ala Asp Leu Tyr Lys Pro Arg Pro Phe
2165 2170 2175

Ser Glu His Ile Asp Ser Asn Ala Leu Ala Gly Tyr Lys Leu Gly
2180 2185 2190

Lys Thr Ala Ser Phe Asn Gly Gly Lys Leu Gly Asp Phe Ala Lys
2195 2200 2205

Pro Ile Ala Pro Ala Arg Val Asn Arg Gly Gly Gly Val Ala Thr
2210 2215 2220

Ala Asp Ile Ala Asn Ile Ser Ala Ser Thr Glu Asn Leu Arg Ser
2225 2230 2235

Glu Ala Arg Ala Arg Ala Arg Leu Lys Ser Asn Thr Glu Leu Gly
2240 2245 2250

Leu Ser Pro Glu Glu Lys Met Gln Leu Ile Arg Ser Arg Leu His
2255 2260 2265

Tyr Asp Gln Asn Arg Ser Leu Lys Pro Lys Gln Leu Glu Glu Met
2270 2275 2280

Pro Ser Gly Asp Leu Ala Ala Arg Ala Arg Lys Met Ser Ala Ser
2285 2290 2295

Lys Ser Val Asn Asp Leu Ala Tyr Met Val Gly Gln Gln Gln Gln
2300 2305 2310

Gln Gln Val Glu Lys Asp Ala Val Leu Gln Ala Lys Ala Ala Asp
2315 2320 2325

Phe Thr Ser Asp Pro Asn Leu Ala Ser Gly Gly Gln Glu Lys Ala
2330 2335 2340

Gly Lys Thr Lys Ser Gly Arg Arg Pro Lys Asp Pro Glu Arg Arg
2345 2350 2355

Lys Ser Leu Ile Gln Ser Leu Ser Ser Phe Phe Gln Lys Gly Ser
2360 2365 2370

Gly Ser Ala Ala Ser Ser Ser Lys Glu Gln Gly Gly Ala Val Ala
2375 2380 2385

Ala Val His Ser Glu Gln Ser Glu Arg Pro Gly Thr Ser Ser Ser
2390 2395 2400

Gly Thr Pro Thr Ile Ser Asp Ala Ala Gly Gly Gly Gly Gly Gly
2405 2410 2415

Gly Gly Val Phe Ser Arg Phe Arg Ile Ser Pro Lys Ser Lys Glu
2420 2425 2430

Lys Ser Lys Ser Cys Phe Asp Leu Arg Asn Phe Gly Phe Gly Asp
2435 2440 2445

Lys Asp Met Leu Val Cys Asn Ala Ala Ser Pro Ala Gly Ala Thr
2450 2455 2460

Ser Ala Ser Gln Lys Asn His Ser Gln Glu Tyr Leu Asn Thr Thr
2465 2470 2475

Asn Asn Ser Arg Tyr Arg Lys Gln Thr Asn Thr Ala Lys Pro Lys
2480 2485 2490

Pro Glu Ser Phe Ser Ser Ser Ser Pro Gln Leu Tyr Ile His Lys
2495 2500 2505

Pro His His Leu Ala Ala Ala His Pro Ser Ala Leu Asp Asp Gln
2510 2515 2520

Thr Pro Pro Pro Ile Pro Pro Leu Pro Leu Asn Tyr Gln Arg Ser
2525 2530 2535

Asp Asp Glu Ser Tyr Ala Asn Glu Thr Arg Glu His Lys Lys Gln
2540 2545 2550

Arg Ala Ile Ser Lys Ala Ser Arg Gln Ala Glu Leu Lys Arg Leu
2555 2560 2565

Arg Ile Ala Gln Glu Ile Gln Arg Glu Gln Glu Glu Ile Glu Val
2570 2575 2580

Gln Leu Lys Asp Leu Glu Ala Arg Gly Val Leu Ile Glu Lys Ala
2585 2590 2595

Leu Arg Gly Glu Ala Gln Asn Ile Glu Asn Leu Asp Ala Thr Lys
2600 2605 2610

Asp Asn Asp Glu Lys Leu Leu Lys Glu Leu Leu Glu Ile Trp Arg
2615 2620 2625

Asn Ile Thr Ala Leu Lys Lys Arg Asp Glu Glu Leu Thr Ile Arg
2630 2635 2640

Gln Gln Glu Leu Gln Leu Glu Tyr Arg His Ala Gln Leu Lys Glu
2645 2650 2655

Glu Leu Asn Leu Arg Leu Ser Cys Asn Lys Leu Asp Lys Ser Ser
2660 2665 2670

Ala Asp Val Ala Ala Glu Gly Ala Ile Leu Asn Glu Met Leu Glu
2675 2680 2685

Ile Val Ala Lys Arg Ala Ala Leu Arg Pro Thr Ala Ser Gln Leu
2690 2695 2700

Asp Leu Thr Ala Ala Gly Ser Ala Ser Thr Ser Ala Glu Ala Thr
2705 2710 2715

Gly Ile Lys Leu Thr Gly Gln Pro His Asp His Glu Glu Ser Ile
2720 2725 2730

Ile

<210> 13
<211> 5411
<212> DNA
<213> Homo sapiens

<400> 13
ggccaagccg gggccccgaa gccagagccg gagccgggcg ggccgcgggg tcatggctgg 60

gccgcggggc gcgctgctgg cctggtgccg ccgccagtgc gagggctacc gcggcgtgga 120
gatccgcgac ctgagcagct ccttccggga cggcctggcc ttctgcgcca tcctgcaccg 180
gcaccggccc gacctgctag attttgattc gctttccaag gacaatgtct tcgagaataa 240
ccgtttggcc tttgaagtgg ctgagaagga gctggggatc cccgctctcc tggaccccaa 300
tgacatggtc tccatgagcg tccctgactg cctcagcatc atgacctatg tgtcccagta 360
ttacaaccac ttctgcagtc ctggccaagc tggtgtctcg ccacccagaa agggccttgc 420
accctgttcc ccgccgtctg tagcacccac tccagtggaa ccagaagatg tggctcaggg 480
cgaggagctc tcctcaggca gcctgtcaga gcagggcacc ggccagaccc ccagcagcac 540
gtgcgcagcc tgccagcagc atgtgcactt ggtgcagcgc tacctggctg acggcaggct 600
gtaccatcgc cactgcttcc ggtgtcggcg gtgctccagc accctgctcc ctggggctta 660
tgagaatggg cctgaggagg gcacctttgt gtgtgcagaa cactgtgcca ggctgggccc 720
ggggacacgg tcggggacca ggcctgggcc cttctcacag ccaaagcagc agcaccagca 780
gcaactcgca gaagatgcca aggatgttcc aggaggcggc cccagctcca gtgctcctgc 840
aggggctgag gccgatggac ccaaggccag ccctgaggcc cggccgcaga tccctaccaa 900
gccccgggtt cctggcaaac tacaggagct ggccagcccc cctgcgggcc gccccacccc 960
tgcccccagg aaggcctctg agagcaccac cccagcaccc cccacgcccc ggccccgctc 1020
cagtctgcag caggagaacc tggtggagca ggctggcagc agcagcctgg tgaacgggag 1080
actgcacgaa ctgcctgtcc ccaagccgag gggacaccg aagccgtccg aggggacacc 1140
agcccccagg aaggaccccc catggatcac gctggtgcag gcagaaccaa agaagaagcc 1200
agccccactt cccccaagca gcagcccggg gccaccaagc caggacagca ggcaggtgga 1260
gaatggaggc accgaggagg tggcccagcc gagcccaacg gccagcctgg agtccaaacc 1320
ctataacccc tttgaggagg aggaggagga caaggaggaa gaggctccag ctgcacccag 1380
cctggccacc agccctgccc tgggccaccc ggagtccaca cccaagtccc tgcacccctg 1440
gtacggcatc acccctacca gcagccccaa gacaaagaag cgccctgccc cgcgcgcacc 1500
cagcgcgtcc ccactggctc tccacgcctc ccgcctctcg cactcggagc cgccctcggc 1560
cacaccatcg ccagcgctca gcgtggagag cctgtcgtct gagagcgcca gccagactgc 1620
aggtgcagag cttctggagc cgccagctgt gcccaagagc tcctcagagc ctgctgtcca 1680
tgcccctggt accccctggaa accctgtcag cctctctacc aactcctccc tggcctcctc 1740
tggggaacta gtggagccta gagtggaaca aatgcctcaa gccagccctg gccttgcccc 1800
caggaccagg ggcagctcag gtccccagcc agccaagccc tgcagtggcg ccaccccaac 1860
gcctctcttg ttggttggag acaggagccc ggtgccttcc cctggaagct cgtccccaca 1920

gctgcaggta aagtcctcct gcaaggagaa tccttttaac cggaagccat cacctgcagc 1980
gtccccagcc acaaagaagg ccaccaaggg atccaagcca gtgaggccac ctgcccctgg 2040
acacggcttt ccactcatca aacgcaaggt ccaggctgac cagtacatcc ctgaggagga 2100
catccatgga gagatggata ccattgagcg ccggctggat gccctggagc accgtggggt 2160
gctgctggag gagaagctgc gtggcggcct gaatgagggc cgtgaggatg acatgctggt 2220
ggactggttc aagctcatcc acgagaagca cctactggtg cggcgagagt ccgagctcat 2280
ctatgtcttc aagcagcaga acctggagca gcgccaggct gatgtcgagt atgagctccg 2340
gtgcctcctc aataagccag aaaaggactg gacggaggag gaccgggccc gggagaaggt 2400
gctgatgcag gagcttgtga ccctcattga gcagcgcaac gctatcatca actgcctgga 2460
tgaggaccgg cagagggagg aagaggaaga caagatgttg gaagccatga tcaagaagaa 2520
agagttccag agggaggctg aacctgaggg caagaagaag gggaagttca agaccatgaa 2580
gatgttgaaa ctgctaggaa acaaacgtga tgccaagagc aagtccccca gagacaagag 2640
ctaacagcac gagaagccag ttggggactg ccccctcctg gagcagctcc tgggctgtgc 2700
tctgtttgaa gggggcgccc tgctcccctc agatcagtca ggaggaagat gactaagggg 2760
agggatcctc tgggtgatgg cctcttcctc ctcagggacc tctgactgct ctgggccaaa 2820
gaatctcttg tttcttctcc gagccccagg cagcggtgat tcagccctgc ccaacctgat 2880
tctgatgact gcggatgctg tgacggaccc aaggggcaaa tagggtccca gggtccaggg 2940
aggggcgcct gctgagcact tccgcccctc accctgccca gcccctgcca tgagctctgg 3000
gctgggtctc cgcctccagg gttctgctct tccaggcagg ccagcaagtg gcgctgggcc 3060
acactggctt cttcctgccc catccctggc tctgagtctc tgtcttcctg tcctgtgcag 3120
gcgcccttgg atctcagttt ccctcactca ggaactctgt ttctgaagtc ttcagttaag 3180
tttgagttta tgactgagtg gcctgtactg tcagacgtga atgggcctga cgggcaaatc 3240
catccctctc tccctcacag ttccaggagc ggcttccctc gtctcccctt actccacagg 3300
gagcctccct tgccaggacc agggctgcga cggccatgct ggggcaggtg agtgctctgt 3360
tagctgctcc cagtgctgtc cccaggctgc agttctggtc cctggttgtc aggtaggaag 3420
ggtgcacttg aagcaggtgc tcatctcggt tccttaacgt ttatagtctg acccctcact 3480
taggctttcc tctgccaccc cggtccaggg aagaggctcg ctcccgccca tggtcatcac 3540
tggtctgtct gctctgttgt ctgttctttc cctgactccc tcccaccgaa ggcctgatgg 3600
ctactcaccc ctctgggatg gctatgggag aggaggagtg atggggaccg ccaccttttc 3660
tgcaggaaat gtgcccagca gctcttggtc aaagcactgt tgctataagc tatctctggg 3720

atgcctctag gccccccttcc ctctacacac ctctgggaaa agattacact gtattaactc 3780
tcgaggagtt tcctcaccaa taaacagaca acctcaactg ccagtgccct gcagcctcgg 3840
gccacagcgg cagccttgtt tgccttccca cctgcctctg ccacacctgg tggctgaaca 3900
tctctggtcg cccagaggcc atgttggggc catcctccaa gagggatctc tgccctcacc 3960
gcctgccact gggcaggatc cctttcctct gcagggagag gtggctcctc ggccatgcag 4020
cccctggcag gctccttcta aacatgcctg ttgacctgga gctggcgcca ccaactccag 4080
ggcctttcca gggccagaca ggtaacacgc atgaacccga gtgacagctc tgacgggctg 4140
tttcggtgtc aggagacaaa gctggcaggg gcaggggtga actggaggca agtcaagtca 4200
cctgtggcct gtggggctga atgtgggccc ggtgttgcca gatcctttgt cataagaagc 4260
tagaaatcca gattttatgt gtgtgtaatt tgtaaatgct gaaagctagc ctgaattttt 4320
tttttttttt tttgagacag agtctcgctc tgtcgcccag gctggagtgc agtggcgcga 4380
tctcagctca ctgcaagctc cgcctcctgg gttcacgcca tcctcctgcc tcggcctcct 4440
gagcagctgg gactacaggc gcatgctacg acgcctggct aatttttttgt atttttagta 4500
gagacggggt ttcaccgtgt taaccaggat ggtctcgatc tcctgacctt gtgatccacc 4560
caccttggcc tcccaaagtg ctgggattac aggcgtgagc caccacgccc ggccactagc 4620
ctgaatttca atcaagggtt ggctgatact gtgtgtccag ggtggactgg atttgtcctg 4680
gggggttctc tggtttgctg cctcctgacc acatgatggg gccttcgagg tcgaggacaa 4740
ctgttcccat tagattgcac cctctgccct caggttcttg agggtgtgtg gacacagagg 4800
ctttccatgg gatgtccctg agccggccct tgattggggc ctcaccattt acagggccgt 4860
tttattctgc aaaccgaaac ttgggtcatg tgacctgatg ggattatggg actccctcca 4920
ggtgcccgag acaaggttga tatttccaaa atattttggt gatttagtgg gacaagcaaa 4980
tgacagaata ccggagaagg cagggatcgt gggtgtcagg agccagaggg gagggggaca 5040
gatgtgctgt gtacaggaca aggtgtcagg tgactccttc ccagcagggc ctcgcagatg 5100
cacaagcacg gagctggtgg gttttgccca agaaaggtca cgcggcacat gcagggattg 5160
gaactcccag gccagggctc taggtcgctc ccaccttttc atgtttcttt ctgtggccat 5220
gggtatagtg gaaagacata aagctaaagc caacttttaa tcctgaatgc actgcttgcc 5280
aggtaaatgc ccttggttgt ggtatcttgt tgagacttag ttttcacaga gggataatga 5340
accgttgcag aggtttattg agatcattaa cagagtggaa ttcagcaccc gccacagcag 5400
ccagccatgg g 5411

<210> 14
<211> 863

<212> PRT
<213> Homo sapiens

<400> 14

Met Ala Gly Pro Arg Gly Ala Leu Leu Ala Trp Cys Arg Arg Gln Cys
1 5 10 15

Glu Gly Tyr Arg Gly Val Glu Ile Arg Asp Leu Ser Ser Ser Phe Arg
20 25 30

Asp Gly Leu Ala Phe Cys Ala Ile Leu His Arg His Arg Pro Asp Leu
35 40 45

Leu Asp Phe Asp Ser Leu Ser Lys Asp Asn Val Phe Glu Asn Asn Arg
50 55 60

Leu Ala Phe Glu Val Ala Glu Lys Glu Leu Gly Ile Pro Ala Leu Leu
65 70 75 80

Asp Pro Asn Asp Met Val Ser Met Ser Val Pro Asp Cys Leu Ser Ile
85 90 95

Met Thr Tyr Val Ser Gln Tyr Tyr Asn His Phe Cys Ser Pro Gly Gln
100 105 110

Ala Gly Val Ser Pro Pro Arg Lys Gly Leu Ala Pro Cys Ser Pro Pro
115 120 125

Ser Val Ala Pro Thr Pro Val Glu Pro Glu Asp Val Ala Gln Gly Glu
130 135 140

Glu Leu Ser Ser Gly Ser Leu Ser Glu Gln Gly Thr Gly Gln Thr Pro
145 150 155 160

Ser Ser Thr Cys Ala Ala Cys Gln Gln His Val His Leu Val Gln Arg
165 170 175

Tyr Leu Ala Asp Gly Arg Leu Tyr His Arg His Cys Phe Arg Cys Arg
180 185 190

Arg Cys Ser Ser Thr Leu Leu Pro Gly Ala Tyr Glu Asn Gly Pro Glu
195 200 205

Glu Gly Thr Phe Val Cys Ala Glu His Cys Ala Arg Leu Gly Pro Gly
210 215 220

Thr Arg Ser Gly Thr Arg Pro Gly Pro Phe Ser Gln Pro Lys Gln Gln
225 230 235 240

His Gln Gln Gln Leu Ala Glu Asp Ala Lys Asp Val Pro Gly Gly Gly
245 250 255

Pro Ser Ser Ser Ala Pro Ala Gly Ala Glu Ala Asp Gly Pro Lys Ala
260 265 270

Ser Pro Glu Ala Arg Pro Gln Ile Pro Thr Lys Pro Arg Val Pro Gly
275 280 285

Lys Leu Gln Glu Leu Ala Ser Pro Pro Ala Gly Arg Pro Thr Pro Ala
290 295 300

Pro Arg Lys Ala Ser Glu Ser Thr Thr Pro Ala Pro Pro Thr Pro Arg
305 310 315 320

Pro Arg Ser Ser Leu Gln Gln Glu Asn Leu Val Glu Gln Ala Gly Ser
325 330 335

Ser Ser Leu Val Asn Gly Arg Leu His Glu Leu Pro Val Pro Lys Pro
340 345 350

Arg Gly Thr Pro Lys Pro Ser Glu Gly Thr Pro Ala Pro Arg Lys Asp
355 360 365

Pro Pro Trp Ile Thr Leu Val Gln Ala Glu Pro Lys Lys Lys Pro Ala
370 375 380

Pro Leu Pro Pro Ser Ser Ser Pro Gly Pro Pro Ser Gln Asp Ser Arg
385 390 395 400

Gln Val Glu Asn Gly Gly Thr Glu Glu Val Ala Gln Pro Ser Pro Thr
405 410 415

Ala Ser Leu Glu Ser Lys Pro Tyr Asn Pro Phe Glu Glu Glu Glu Glu
420 425 430

Asp Lys Glu Glu Glu Ala Pro Ala Ala Pro Ser Leu Ala Thr Ser Pro
435 440 445

Ala Leu Gly His Pro Glu Ser Thr Pro Lys Ser Leu His Pro Trp Tyr
450 455 460

Gly Ile Thr Pro Thr Ser Ser Pro Lys Thr Lys Lys Arg Pro Ala Pro

465 470 475 480

Arg Ala Pro Ser Ala Ser Pro Leu Ala Leu His Ala Ser Arg Leu Ser
485 490 495

His Ser Glu Pro Pro Ser Ala Thr Pro Ser Pro Ala Leu Ser Val Glu
500 505 510

Ser Leu Ser Ser Glu Ser Ala Ser Gln Thr Ala Gly Ala Glu Leu Leu
515 520 525

Glu Pro Pro Ala Val Pro Lys Ser Ser Ser Glu Pro Ala Val His Ala
530 535 540

Pro Gly Thr Pro Gly Asn Pro Val Ser Leu Ser Thr Asn Ser Ser Leu
545 550 555 560

Ala Ser Ser Gly Glu Leu Val Glu Pro Arg Val Glu Gln Met Pro Gln
565 570 575

Ala Ser Pro Gly Leu Ala Pro Arg Thr Arg Gly Ser Ser Gly Pro Gln
580 585 590

Pro Ala Lys Pro Cys Ser Gly Ala Thr Pro Thr Pro Leu Leu Leu Val
595 600 605

Gly Asp Arg Ser Pro Val Pro Ser Pro Gly Ser Ser Ser Pro Gln Leu
610 615 620

Gln Val Lys Ser Ser Cys Lys Glu Asn Pro Phe Asn Arg Lys Pro Ser
625 630 635 640

Pro Ala Ala Ser Pro Ala Thr Lys Lys Ala Thr Lys Gly Ser Lys Pro
645 650 655

Val Arg Pro Pro Ala Pro Gly His Gly Phe Pro Leu Ile Lys Arg Lys
660 665 670

Val Gln Ala Asp Gln Tyr Ile Pro Glu Glu Asp Ile His Gly Glu Met
675 680 685

Asp Thr Ile Glu Arg Arg Leu Asp Ala Leu Glu His Arg Gly Val Leu
690 695 700

Leu Glu Glu Lys Leu Arg Gly Gly Leu Asn Glu Gly Arg Glu Asp Asp
705 710 715 720

Met Leu Val Asp Trp Phe Lys Leu Ile His Glu Lys His Leu Leu Val
725 730 735

Arg Arg Glu Ser Glu Leu Ile Tyr Val Phe Lys Gln Gln Asn Leu Glu
740 745 750

Gln Arg Gln Ala Asp Val Glu Tyr Glu Leu Arg Cys Leu Leu Asn Lys
755 760 765

Pro Glu Lys Asp Trp Thr Glu Glu Asp Arg Ala Arg Glu Lys Val Leu
770 775 780

Met Gln Glu Leu Val Thr Leu Ile Glu Gln Arg Asn Ala Ile Ile Asn
785 790 795 800

Cys Leu Asp Glu Asp Arg Gln Arg Glu Glu Glu Glu Asp Lys Met Leu
805 810 815

Glu Ala Met Ile Lys Lys Lys Glu Phe Gln Arg Glu Ala Glu Pro Glu
820 825 830

Gly Lys Lys Lys Gly Lys Phe Lys Thr Met Lys Met Leu Lys Leu Leu
835 840 845

Gly Asn Lys Arg Asp Ala Lys Ser Lys Ser Pro Arg Asp Lys Ser
850 855 860

<210> 15
<211> 2685
<212> DNA
<213> Homo sapiens

<400> 15
atggcggcca tcagggcgct gcaacagtgg tgccggcagc agtgcgaggg ctaccgcgac 60

gtgaatatct gcaacatgac cacgtcgttc cgcgacggcc tggctttctg cgccatcctg 120

caccgccacc ggcccgacct cataaacttc agtgctctca agaaggaaaa tatttatgaa 180

aacaataaac tggccttccg cgtggccgag gagcacttgg gcatcccagc cttgctggat 240

gccgaggaca tggtggcctt gaaggtgcct gaccggctga gcatcttgac ctacgtgtcc 300

cagtattaca actacttcca cggccgctcc cccattgggg gcatggcagg cgtgaagagg 360

gcctcggagg actctgagga ggagccgtca gggaagaagg ctccagtcca ggcggccaag 420

ctgccctcgc ccgccccagc ccggaagcct ccactatctc cagcccagac aaaccctgtg 480

gtccagagga ggaatgaggg tgcagggggc ccgccccccca agactgacca ggcattggcg 540

ggcagcttgg tcagcagcac ctgcggggtc tgcggcaagc acgtgcacct ggtacagcgg 600
cacctggccg acgggaggct ttaccaccgg agctgcttca ggtgtaagca gtgctcctgc 660
acgctgcact cggggggccta caaggccaca ggagagccgg gcaccttcgt ctgcaccagc 720
cacctccccg cagccgcctc tgcaagcccc aagttgacgg gtctggtccc ccgacagcca 780
ggggccatgg gtgtggattc caggacctcc tgttccccac agaaggccca ggaggcaaac 840
aaggccagac cgttggcctg ggagcctgct gcgggcaact cgcctgccag ggcttccgtt 900
ccagctgcac ccaaccctgc agccaccagc gccacgtccg tccacgtgag gagcccagcc 960
aggccctctg agagccgcct ggcccccact cccacggagg ggaaagtccg ccctcgtgtg 1020
accaatagct ccccgatggg ctggtcgtca gctgccccgt gcacagcagc ggctgcctcc 1080
catcccgccg tgcccccgag tgccccagac cctcgcccgg ccacacccca gggcggggga 1140
gccccccgag tggcagctcc tcaaaccaca ctcagttcaa gctccacatc tgcagccacg 1200
gtggaccccc cagcctggac cccgtccgcc tccaggaccc agcaggcccg gaataagttt 1260
ttccaaacat cagcagtgcc ccccggcacc agcctttctg gcagaggtcc caccccgtca 1320
cttgttctat ccaaggacag cagcaaggag caggcgcgga acttcctcaa gcaggccctc 1380
tcagcgctgg aagaggctgg cgctccggcg cctggcaggc cctccccagc cactgccgct 1440
gttcccagtt ctcagcccaa aactgaagca ccacaagcaa gtcccttagc caagccgtta 1500
cagtcctcgt ctccccgggt gcttggcctc ccttcgagga tggaaccgcc agccccgctg 1560
agcacgagca gtacctctca ggcatccgcg ttgcccccgg caggcaggag gaacttggct 1620
gaatcctcag gggtcggcag ggtgggtgct ggctccaggc cgaagccaga ggccccgatg 1680
gcaaagggta aaagcaccac cttaacgcag gacatgagca ccagcctcca ggaaggccag 1740
gaggacgggc cggcaggatg gagagcgaat ctgaagcccg tggacaggag gagcccagct 1800
gagaggactc tgaagcccaa ggaaccacgg gccctggcag agccgagggc gggggaggcc 1860
cccaggaagg tctcaggcag ctttgctggg agtgtccaca tcaccctgac ccccgtgagg 1920
cctgacagga ccccacgccc agccagccca ggacccagcc tcccagccag gtccccctcc 1980
ccaccccgcc gcaggagact ggccgtccct gccagcctcg acgtttgtga caactggctt 2040
cggccggagc cccctggcca ggaagcccga gtgcagagct ggaaggagga ggagaagaaa 2100
cctcaccttc agggcagacc agggagaccc ttgtccccgg ccaatgtccc tgctctgcct 2160
ggcgagacgg tgacctcccc agtcaggctg caccccgact acctctcccc ggaggagata 2220
cagaggcagc tgcaggacat cgagaggcgg ctggacgccc tggagctccg cggcgtggag 2280
ctggagaagc gactgcgggc ggccgaggga gatgacgctg aggatagcct catggtggac 2340

tggttctggc tcattcacga gaagcagctt ctgctgagac aggagtcaga gctgatgtac 2400

aagtccaagg cccagcgtct ggaggagcag cagctggaca tcgagggcga gctgcgccgg 2460

ctcatggcca agcccgaggc tctgaagtca ctgcaggagc ggcggcggga gcaggagctg 2520

ctggagcagt acgtgagcac cgtgaacgac cgcagtgaca tcgtggactc gctggacgag 2580

gaccggctcc gggaacaaga ggaggatcag atgctgcggg acatgattga gaagctgggc 2640

ctccagagga agaagtccaa gttccgcttg tccaagatct ggtca 2685

<210> 16
<211> 904
<212> PRT
<213> Homo sapiens

<400> 16

Met Ala Ala Ile Arg Ala Leu Gln Gln Trp Cys Arg Gln Gln Cys Glu
1 5 10 15

Gly Tyr Arg Asp Val Asn Ile Cys Asn Met Thr Thr Ser Phe Arg Asp
20 25 30

Gly Leu Ala Phe Cys Ala Ile Leu His Arg His Arg Pro Asp Leu Ile
35 40 45

Asn Phe Ser Ala Leu Lys Lys Glu Asn Ile Tyr Glu Asn Asn Lys Leu
50 55 60

Ala Phe Arg Val Ala Glu Glu His Leu Gly Ile Pro Ala Leu Leu Asp
65 70 75 80

Ala Glu Asp Met Val Ala Leu Lys Val Pro Asp Arg Leu Ser Ile Leu
85 90 95

Thr Tyr Val Ser Gln Tyr Tyr Asn Tyr Phe His Gly Arg Ser Pro Ile
100 105 110

Gly Gly Met Ala Gly Val Lys Arg Ala Ser Glu Asp Ser Glu Glu Glu
115 120 125

Pro Ser Gly Lys Lys Ala Pro Val Gln Ala Ala Lys Leu Pro Ser Pro
130 135 140

Ala Pro Ala Arg Lys Pro Pro Leu Ser Pro Ala Gln Thr Asn Pro Val
145 150 155 160

Val Gln Arg Arg Asn Glu Gly Ala Gly Gly Pro Pro Pro Lys Thr Asp

165 170 175

Gln Ala Leu Ala Gly Ser Leu Val Ser Ser Thr Cys Gly Val Cys Gly
180 185 190

Lys His Val His Leu Val Gln Arg His Leu Ala Asp Gly Arg Leu Tyr
195 200 205

His Arg Ser Cys Phe Arg Cys Lys Gln Cys Ser Cys Thr Leu His Ser
210 215 220

Gly Ala Tyr Lys Ala Thr Gly Glu Pro Gly Thr Phe Val Cys Thr Ser
225 230 235 240

His Leu Pro Ala Ala Ala Ser Ala Ser Pro Lys Leu Thr Gly Leu Val
245 250 255

Pro Arg Gln Pro Gly Ala Met Gly Val Asp Ser Arg Thr Ser Cys Ser
260 265 270

Pro Gln Lys Ala Gln Glu Ala Asn Lys Ala Arg Pro Leu Ala Trp Glu
275 280 285

Pro Ala Ala Gly Asn Ser Pro Ala Arg Ala Ser Val Pro Ala Ala Pro
290 295 300

Asn Pro Ala Ala Thr Ser Ala Thr Ser Val His Val Arg Ser Pro Ala
305 310 315 320

Arg Pro Ser Glu Ser Arg Leu Ala Pro Thr Pro Thr Glu Gly Lys Val
325 330 335

Arg Pro Arg Val Thr Asn Ser Ser Pro Met Gly Trp Ser Ser Ala Ala
340 345 350

Pro Cys Thr Ala Ala Ala Ala Ser His Pro Ala Val Pro Pro Ser Ala
355 360 365

Pro Asp Pro Arg Pro Ala Thr Pro Gln Gly Gly Gly Ala Pro Arg Val
370 375 380

Ala Ala Pro Gln Thr Thr Leu Ser Ser Ser Ser Thr Ser Ala Ala Thr
385 390 395 400

Val Asp Pro Pro Ala Trp Thr Pro Ser Ala Ser Arg Thr Gln Gln Ala
405 410 415

Arg Asn Lys Phe Phe Gln Thr Ser Ala Val Pro Pro Gly Thr Ser Leu
420 425 430

Ser Gly Arg Gly Pro Thr Pro Ser Leu Val Leu Ser Lys Asp Ser Ser
435 440 445

Lys Glu Gln Ala Arg Asn Phe Leu Lys Gln Ala Leu Ser Ala Leu Glu
450 455 460

Glu Ala Gly Ala Pro Ala Pro Gly Arg Pro Ser Pro Ala Thr Ala Ala
465 470 475 480

Val Pro Ser Ser Gln Pro Lys Thr Glu Ala Pro Gln Ala Ser Pro Leu
485 490 495

Ala Lys Pro Leu Gln Ser Ser Ser Pro Arg Val Leu Gly Leu Pro Ser
500 505 510

Arg Met Glu Pro Pro Ala Pro Leu Ser Thr Ser Ser Thr Ser Gln Ala
515 520 525

Ser Ala Leu Pro Pro Ala Gly Arg Arg Asn Leu Ala Glu Ser Ser Gly
530 535 540

Val Gly Arg Val Gly Ala Gly Ser Arg Pro Lys Pro Glu Ala Pro Met
545 550 555 560

Ala Lys Gly Lys Ser Thr Thr Leu Thr Gln Asp Met Ser Thr Ser Leu
565 570 575

Gln Glu Gly Gln Glu Asp Gly Pro Ala Gly Trp Arg Ala Asn Leu Lys
580 585 590

Pro Val Asp Arg Arg Ser Pro Ala Glu Arg Thr Leu Lys Pro Lys Glu
595 600 605

Pro Arg Ala Leu Ala Glu Pro Arg Ala Gly Glu Ala Pro Arg Lys Val
610 615 620

Ser Gly Ser Phe Ala Gly Ser Val His Ile Thr Leu Thr Pro Val Arg
625 630 635 640

Pro Asp Arg Thr Pro Arg Pro Ala Ser Pro Gly Pro Ser Leu Pro Ala
645 650 655

Arg Ser Pro Ser Pro Pro Arg Arg Arg Arg Leu Ala Val Pro Ala Ser
660 665 670

Leu Asp Val Cys Asp Asn Trp Leu Arg Pro Glu Pro Pro Gly Gln Glu
675 680 685

Ala Arg Val Gln Ser Trp Lys Glu Glu Glu Lys Lys Pro His Leu Gln
690 695 700

Gly Arg Pro Gly Arg Pro Leu Ser Pro Ala Asn Val Pro Ala Leu Pro
705 710 715 720

Gly Glu Thr Val Thr Ser Pro Val Arg Leu His Pro Asp Tyr Leu Ser
725 730 735

Pro Glu Glu Ile Gln Arg Gln Leu Gln Asp Ile Glu Arg Arg Leu Asp
740 745 750

Ala Leu Glu Leu Arg Gly Val Glu Leu Glu Lys Arg Leu Arg Ala Ala
755 760 765

Glu Gly Asp Asp Ala Glu Asp Ser Leu Met Val Asp Trp Phe Trp Leu
770 775 780

Ile His Glu Lys Gln Leu Leu Leu Arg Gln Glu Ser Glu Leu Met Tyr
785 790 795 800

Lys Ser Lys Ala Gln Arg Leu Glu Glu Gln Gln Leu Asp Ile Glu Gly
805 810 815

Glu Leu Arg Arg Leu Met Ala Lys Pro Glu Ala Leu Lys Ser Leu Gln
820 825 830

Glu Arg Arg Arg Glu Gln Glu Leu Leu Glu Gln Tyr Val Ser Thr Val
835 840 845

Asn Asp Arg Ser Asp Ile Val Asp Ser Leu Asp Glu Asp Arg Leu Arg
850 855 860

Glu Gln Glu Glu Asp Gln Met Leu Arg Asp Met Ile Glu Lys Leu Gly
865 870 875 880

Leu Gln Arg Lys Lys Ser Lys Phe Arg Leu Ser Lys Ile Trp Ser Pro
885 890 895

Lys Ser Lys Ser Ser Pro Ser Gln
900

<210> 17
<211> 3351
<212> DNA
<213> Drosophila

<400> 17
caccgttttg ccttgtttac atttaatgta atttgtacca gataaatctt atttgatttt 60

actcaagtca agtggcacga tgtgcttttg ttgttcactt agtttacagt ttgtgactgc 120

tattttcgtt tagttatatt cgccaataag ccagcaatta acccactaaa atgtctgatc 180

gtcgtggcac aaaagttgga actggtacga aggctttgga gtattggtgc cgagttgtga 240

cccaaggata taatggggtc aaggtggaga acatgaccac ttcctggcga aatggacttg 300

ccttctgcgc cataatacac cactttcggc cggatcttat agatttcgac cgacttaaag 360

cagacgatat ctatgagaac aacgatttgg cctttacaac ggccgagaaa tatttgggaa 420

ttcccgcact gctcgatgca gctgacatgg tttcgtatga agtacctgat agactctcca 480

tactaacgta tttatcccag ttttacaagg tgctcggcaa gagcctgaag cacccgaagc 540

cagaggagcc acttggcgaa gagagtgagc caccgcaaaa ggtgatgcac attgttggga 600

tgcctcgacg tgataagtgc cagaaatgca accttcctgt tttctcgcc gaaagggtcc 660

ttgtcggcaa acgggcttat catcgcacat gcttaaaatg tgcaagatgt tcgtcattgc 720

tcacacccgg gagtttttat gaaacggagg tcaataacat atactgttgc gagacttgtc 780

ccgatgaaga aagtgaaccg gaatctgaca ttttaaagtt aaagacaacc actactgatt 840

ctccgaatga taaacaaatg gtggcacaaa gttctgatta ctctgaagct gaagataaac 900

aagaagacct ggaagataat gatatacgta ctactgataa gcctgaaaac ttccaaccgc 960

cgtccaacaa agatgaacaa aataatgaac taactattaa tccggtaaac cctatattat 1020

cagaggagcg taaaatctct tttattccac tagacgaaga agatggtggc cttattgaac 1080

aatataataa atcaacaact cctgtaaagc ccgctatacc agagaaacca aaagtctcta 1140

ctcttccact agacgatgaa caacatgctg gcgttgaaca aaacaatgat ctggcggtta 1200

gtccggaaaa tgacattcca aaagagaaac ttaagatttc ttcggtctca atatatcttg 1260

aagatgaccg ccttgttgta gatgcaatac atccggataa tttagataaa caagaagctt 1320

tgaacaatac atcagacgca ctcattccag aatcacagga agcaccgatt cctgaaaata 1380

acactcaagt agccattaaa ccagaagacc atataagccc gcgtaaagaa aataaaattt 1440

tctcaaacac agaaagctgc tctaagcagg aaggtgttct cccaaaacaa atggatctcg 1500

agtctcccaa ggacaaggta attgagacaa aagcatctga aactgactat ccagaggatc 1560

taaatccctt taaagatgat gactcatcta aaggtgccaa cccattcgac agctctgatg 1620
atgaagtgga gcttcttaaa gccattccag cccaacaaag caaaggaaaa gttgtaccgc 1680
cccgtccacc gccacctaaa attggccttt catccatctc caatccgtca gagaaaccac 1740
attccagccc tacactttcg catgggaaaa aaatgccgat gccaacaccg agaatatcta 1800
tatcaaaaac ccagactcca gcaaaaccaa tgacgcatca aggccagaaa agtagtattt 1860
cctcttcgtc ttcggagcat ttgaatagca taagaacgtt cgatcgcgga gcagatgacc 1920
gtggctcaag catctccttg ccatctgcaa atgggcctcg caagccactt cgcgcctcag 1980
tggggagtcc gctgcgaagc gaagaatcga gtcccactac cagtcttagc tctattacct 2040
ctccgatgcg gaagaagcgc caggcacctt tgcccccaat acaaacggac tttgacagtg 2100
atcctggatt ttcaaaattg tccgacgaac aaaaggcatt gttgcacact cagcttaagg 2160
ccccaaatct gggtgattca accagaagac taataccact tgaccagagt ctgctatccg 2220
atgaggcaac cgagtcgagt aattacgacg agagcttaag tacatccaat gccgacgagg 2280
aggtaaatgt ggtataccga cgcattttgg tgccgccaac tcaacccgaa aacactgttg 2340
aacgtagcaa ggaggatcaa aaatcgccta tcgtgtataa cgacttcgat agaaacgtaa 2400
gcccattggg gcacaataaa tccactcatg ggaaatggaa gaggcgaaag ggaccagcac 2460
cagcggttcc aataccaccg cgcaaagtct tacaaaggct tccgctgcaa gaaattcgcc 2520
acgagttcga aattattgcg gtacagcagc tgggtcttga gaaacaaggc gttattctgg 2580
agaaaatgat cagggatcgc tgtgagcgtt ccttagatgc caccgatact gatggtcccg 2640
aatccgcaga agtgctaacc aactcaaagg aagtggaaga tcttatattg caactgttcg 2700
agctggtcaa cgagaaaaac gaattgttcc gcagacaagc ggaactgatg taccttcgac 2760
gtcagcatcg cctggagcag gagcaggcgg acatagagca tgagatacga gtattaatgg 2820
ggcaaccgga gcacaacaaa accgattcgg ataaagccca tgaggaagta ttaatcaatc 2880
gccttgtgaa ggtggttgaa atgcgtaacg aagtaattga tagcctagag actgaccgag 2940
ttcgcgaggc gcgagaggat atgagcatca agaatcggct tcatatatac aattctgagc 3000
gcgaggaacc accagcacat ccgagaagcg ctgacaaatc atccaaaaag ctgtctaaaa 3060
aggaacgaaa aaaacttaag gaggaaaata agctgggcaa gggcaaaaag tcggacctcg 3120
acaaggacgt cgatgagtcc gaacaggcac cagctttgga aaaggttaaa aagaagcgaa 3180
acttgttctt tttgaaaatg tagttggggg ctggttgagt tgaagtgttt acgcatcgaa 3240
caattgaaat ttattttata taaatttgta acctattaag tttctattgc catatgtgag 3300
catttcaata aataaagtat atattttcct ttaaaaaaaa aaaaaaaaaa a 3351

<210> 18
<211> 1010
<212> PRT
<213> Drosophila

<400> 18

Met Ser Asp Arg Arg Gly Thr Lys Val Gly Thr Gly Thr Lys Ala Leu
1 5 10 15

Glu Tyr Trp Cys Arg Val Val Thr Gln Gly Tyr Asn Gly Val Lys Val
20 25 30

Glu Asn Met Thr Thr Ser Trp Arg Asn Gly Leu Ala Phe Cys Ala Ile
35 40 45

Ile His His Phe Arg Pro Asp Leu Ile Asp Phe Asp Arg Leu Lys Ala
50 55 60

Asp Asp Ile Tyr Glu Asn Asn Asp Leu Ala Phe Thr Thr Ala Glu Lys
65 70 75 80

Tyr Leu Gly Ile Pro Ala Leu Leu Asp Ala Ala Asp Met Val Ser Tyr
85 90 95

Glu Val Pro Asp Arg Leu Ser Ile Leu Thr Tyr Leu Ser Gln Phe Tyr
100 105 110

Lys Val Leu Gly Lys Ser Leu Lys His Pro Lys Pro Glu Glu Pro Leu
115 120 125

Gly Glu Glu Ser Glu Pro Pro Gln Lys Val Met His Ile Val Gly Met
130 135 140

Pro Arg Arg Asp Lys Cys Gln Lys Cys Asn Leu Pro Val Phe Leu Ala
145 150 155 160

Glu Arg Val Leu Val Gly Lys Arg Ala Tyr His Arg Thr Cys Leu Lys
165 170 175

Cys Ala Arg Cys Ser Ser Leu Leu Thr Pro Gly Ser Phe Tyr Glu Thr
180 185 190

Glu Val Asn Asn Ile Tyr Cys Cys Glu Thr Cys Pro Asp Glu Glu Ser
195 200 205

Glu Pro Glu Ser Asp Ile Leu Lys Leu Lys Thr Thr Thr Thr Asp Ser

210 215 220

Pro Asn Asp Lys Gln Met Val Ala Gln Ser Ser Asp Tyr Ser Glu Ala
225 230 235 240

Glu Asp Lys Gln Glu Asp Leu Glu Asp Asn Asp Ile Arg Thr Thr Asp
245 250 255

Lys Pro Glu Asn Phe Gln Pro Pro Ser Asn Lys Asp Glu Gln Asn Asn
260 265 270

Glu Leu Thr Ile Asn Pro Val Asn Pro Ile Leu Ser Glu Glu Arg Lys
275 280 285

Ile Ser Phe Ile Pro Leu Asp Glu Glu Asp Gly Gly Leu Ile Glu Gln
290 295 300

Tyr Asn Lys Ser Thr Thr Pro Val Lys Pro Ala Ile Pro Glu Lys Pro
305 310 315 320

Lys Val Ser Thr Leu Pro Leu Asp Asp Glu Gln His Ala Gly Val Glu
325 330 335

Gln Asn Asn Asp Leu Ala Val Ser Pro Glu Asn Asp Ile Pro Lys Glu
340 345 350

Lys Leu Lys Ile Ser Ser Val Ser Ile Tyr Leu Glu Asp Asp Arg Leu
355 360 365

Val Val Asp Ala Ile His Pro Asp Asn Leu Asp Lys Gln Glu Ala Leu
370 375 380

Asn Asn Thr Ser Asp Ala Leu Ile Pro Glu Ser Gln Glu Ala Pro Ile
385 390 395 400

Pro Glu Asn Asn Thr Gln Val Ala Ile Lys Pro Glu Asp His Ile Ser
405 410 415

Pro Arg Lys Glu Asn Lys Ile Phe Ser Asn Thr Glu Ser Cys Ser Lys
420 425 430

Gln Glu Gly Val Leu Pro Lys Gln Met Asp Leu Glu Ser Pro Lys Asp
435 440 445

Lys Val Ile Glu Thr Lys Ala Ser Glu Thr Asp Tyr Pro Glu Asp Leu
450 455 460

Asn Pro Phe Lys Asp Asp Asp Ser Ser Lys Gly Ala Asn Pro Phe Asp
465 470 475 480

Ser Ser Asp Asp Glu Val Glu Leu Leu Lys Ala Ile Pro Ala Gln Gln
485 490 495

Ser Lys Gly Lys Val Val Pro Pro Arg Pro Pro Pro Pro Lys Ile Gly
500 505 510

Leu Ser Ser Ile Ser Asn Pro Ser Glu Lys Pro His Ser Ser Pro Thr
515 520 525

Leu Ser His Gly Lys Lys Met Pro Met Pro Thr Pro Arg Ile Ser Ile
530 535 540

Ser Lys Thr Gln Thr Pro Ala Lys Pro Met Thr His Gln Gly Gln Lys
545 550 555 560

Ser Ser Ile Ser Ser Ser Ser Ser Glu His Leu Asn Ser Ile Arg Thr
565 570 575

Phe Asp Arg Gly Ala Asp Asp Arg Gly Ser Ser Ile Ser Leu Pro Ser
580 585 590

Ala Asn Gly Pro Arg Lys Pro Leu Arg Ala Ser Val Gly Ser Pro Leu
595 600 605

Arg Ser Glu Glu Ser Ser Pro Thr Thr Ser Leu Ser Ser Ile Thr Ser
610 615 620

Pro Met Arg Lys Lys Arg Gln Ala Pro Leu Pro Pro Ile Gln Thr Asp
625 630 635 640

Phe Asp Ser Asp Pro Gly Phe Ser Lys Leu Ser Asp Glu Gln Lys Ala
645 650 655

Leu Leu His Thr Gln Leu Lys Ala Pro Asn Leu Gly Asp Ser Thr Arg
660 665 670

Arg Leu Ile Pro Leu Asp Gln Ser Leu Leu Ser Asp Glu Ala Thr Glu
675 680 685

Ser Ser Asn Tyr Asp Glu Ser Leu Ser Thr Ser Asn Ala Asp Glu Glu
690 695 700

Val Asn Val Val Tyr Arg Arg Ile Leu Val Pro Pro Thr Gln Pro Glu
705 710 715 720

Asn Thr Val Glu Arg Ser Lys Glu Asp Gln Lys Ser Pro Ile Val Tyr
725 730 735

Asn Asp Phe Asp Arg Asn Val Ser Pro Leu Gly His Asn Lys Ser Thr
740 745 750

His Gly Lys Trp Lys Arg Arg Lys Gly Pro Ala Pro Ala Val Pro Ile
755 760 765

Pro Pro Arg Lys Val Leu Gln Arg Leu Pro Leu Gln Glu Ile Arg His
770 775 780

Glu Phe Glu Ile Ile Ala Val Gln Gln Leu Gly Leu Glu Lys Gln Gly
785 790 795 800

Val Ile Leu Glu Lys Met Ile Arg Asp Arg Cys Glu Arg Ser Leu Asp
805 810 815

Ala Thr Asp Thr Asp Gly Pro Glu Ser Ala Glu Val Leu Thr Asn Ser
820 825 830

Lys Glu Val Glu Asp Leu Ile Leu Gln Leu Phe Glu Leu Val Asn Glu
835 840 845

Lys Asn Glu Leu Phe Arg Arg Gln Ala Glu Leu Met Tyr Leu Arg Arg
850 855 860

Gln His Arg Leu Glu Gln Glu Gln Ala Asp Ile Glu His Glu Ile Arg
865 870 875 880

Val Leu Met Gly Gln Pro Glu His Asn Lys Thr Asp Ser Asp Lys Ala
885 890 895

His Glu Glu Val Leu Ile Asn Arg Leu Val Lys Val Val Glu Met Arg
900 905 910

Asn Glu Val Ile Asp Ser Leu Glu Thr Asp Arg Val Arg Glu Ala Arg
915 920 925

Glu Asp Met Ser Ile Lys Asn Arg Leu His Ile Tyr Asn Ser Glu Arg
930 935 940

Glu Glu Pro Pro Ala His Pro Arg Ser Ala Asp Lys Ser Ser Lys Lys
945 950 955 960

Leu Ser Lys Lys Glu Arg Lys Lys Leu Lys Glu Glu Asn Lys Leu Gly
965 970 975

Lys Gly Lys Lys Ser Asp Leu Asp Lys Asp Val Asp Glu Ser Glu Gln
980 985 990

Ala Pro Ala Leu Glu Lys Val Lys Lys Lys Arg Asn Leu Phe Phe Leu
995 1000 1005

Lys Met
1010

<210> 19
<211> 202
<212> PRT
<213> Artificial sequence

<220>
<223> Drosophila truncated mutant

<400> 19

Met Asn Tyr Gln Arg Ser Asp Asp Glu Ser Tyr Ala Asn Glu Thr Arg
1 5 10 15

Glu His Lys Lys Gln Arg Ala Ile Ser Lys Ala Ser Arg Gln Ala Glu
20 25 30

Leu Lys Arg Leu Arg Ile Ala Gln Glu Ile Gln Arg Glu Gln Glu Glu
35 40 45

Ile Glu Val Gln Leu Lys Asp Leu Glu Ala Arg Gly Val Leu Ile Glu
50 55 60

Lys Ala Leu Arg Gly Glu Ala Gln Asn Ile Glu Asn Leu Asp Ala Thr
65 70 75 80

Lys Asp Asn Asp Glu Lys Leu Leu Lys Glu Leu Leu Glu Ile Trp Arg
85 90 95

Asn Ile Thr Ala Leu Lys Lys Arg Asp Glu Glu Leu Thr Ile Arg Gln
100 105 110

Gln Glu Leu Gln Leu Glu Tyr Arg His Ala Gln Leu Lys Glu Glu Leu
115 120 125

Asn Leu Arg Leu Ser Cys Asn Lys Leu Asp Lys Ser Ser Ala Asp Val
130 135 140

Ala Ala Glu Gly Ala Ile Leu Asn Glu Met Leu Glu Ile Val Ala Lys
145 150 155 160

Arg Ala Ala Leu Arg Pro Thr Ala Ser Gln Leu Asp Leu Thr Ala Ala
165 170 175

Gly Ser Ala Ser Thr Ser Ala Glu Ala Thr Gly Ile Lys Leu Thr Gly
180 185 190

Gln Pro His Asp His Glu Glu Ser Ile Ile
195 200

<210> 20
<211> 3002
<212> PRT
<213> Artificial sequence

<220>
<223> Drosophila G-W mutant. G residues 134, 136, 139 of Drosophila MI CAL changed to W residues

<400> 20

Met Ser Arg Gln His Gln Arg His His Gln Gln His His His Leu Pro
1 5 10 15

Pro His Gln Gln Pro Gln Gln Gln Met Pro Gln Gln Gln Gln Gln Leu
20 25 30

Thr Ala Gln Gln Gln Gln Gln Gln Gln Leu Leu Met Ala Glu His Ala
35 40 45

Ala Ala Ala Glu Ala Ala Glu Leu Phe Asp Leu Leu Cys Val Ala Thr
50 55 60

Thr Met Arg Gln Ile Leu Ala Leu His Arg Ala Met Cys Glu Ala Val
65 70 75 80

Gly Leu Arg Pro Ser Pro Leu Asn Asp Phe Tyr Pro Arg Leu Lys Ala
85 90 95

Lys Val Arg Ser Trp Lys Ala Gln Ala Leu Trp Lys Lys Phe Asp Ala
100 105 110

Arg Ala Ala His Arg Val Tyr Gly Lys Gly Ala Ala Cys Thr Gly Thr

```
        115                 120                 125

Arg Val Leu Val Ile Trp Ala Trp Pro Cys Trp Leu Arg Thr Ala Ile
    130                 135                 140

Glu Ala Gln Leu Leu Gly Ala Lys Val Val Val Leu Glu Lys Arg Asp
145                 150                 155                 160

Arg Ile Thr Arg Asn Asn Val Leu His Leu Trp Pro Phe Val Ile Thr
                165                 170                 175

Asp Leu Arg Asn Leu Gly Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala
            180                 185                 190

Gly Ser Ile Asp His Ile Ser Ile Arg Gln Leu Gln Cys Met Leu Leu
        195                 200                 205

Lys Val Ala Leu Leu Leu Gly Val Glu Ile His Glu Gly Val Ser Phe
    210                 215                 220

Asp His Ala Val Glu Pro Ser Gly Asp Gly Gly Gly Trp Arg Ala Ala
225                 230                 235                 240

Val Thr Pro Ala Asp His Pro Val Ser His Tyr Glu Phe Asp Val Leu
                245                 250                 255

Ile Gly Ala Asp Gly Lys Arg Asn Met Leu Asp Phe Arg Arg Lys Glu
            260                 265                 270

Phe Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Lys
        275                 280                 285

Lys Thr Glu Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe
    290                 295                 300

Ile Phe Asn Gln Ala Phe Phe Lys Glu Leu Tyr Gly Lys Thr Gly Ile
305                 310                 315                 320

Asp Leu Glu Asn Ile Val Tyr Tyr Lys Asp Glu Thr His Tyr Phe Val
                325                 330                 335

Met Thr Ala Lys Lys His Ser Leu Ile Asp Lys Gly Val Ile Ile Glu
            340                 345                 350

Asp Met Ala Asp Pro Gly Glu Leu Leu Ala Pro Ala Asn Val Asp Thr
        355                 360                 365
```

Gln Lys Leu His Asp Tyr Ala Arg Glu Ala Ala Glu Phe Ser Thr Gln
370 375 380

Tyr Gln Met Pro Asn Leu Glu Phe Ala Val Asn His Tyr Gly Lys Pro
385 390 395 400

Asp Val Ala Met Phe Asp Phe Thr Ser Met Phe Ala Ala Glu Met Ser
405 410 415

Cys Arg Val Ile Val Arg Lys Gly Ala Arg Leu Met Gln Cys Leu Val
420 425 430

Gly Asp Ser Leu Leu Glu Pro Phe Trp Pro Thr Gly Ser Gly Cys Ala
435 440 445

Arg Gly Phe Leu Ser Ser Met Asp Ala Ala Tyr Ala Ile Lys Leu Trp
450 455 460

Ser Asn Pro Gln Asn Ser Thr Leu Gly Val Leu Ala Gln Arg Glu Ser
465 470 475 480

Ile Tyr Arg Leu Leu Asn Gln Thr Thr Pro Asp Thr Leu Gln Arg Asp
485 490 495

Ile Ser Ala Tyr Thr Val Asp Pro Ala Thr Arg Tyr Pro Asn Leu Asn
500 505 510

Arg Glu Ser Val Asn Ser Trp Gln Val Lys His Leu Val Asp Thr Asp
515 520 525

Asp Pro Ser Ile Leu Glu Gln Thr Phe Met Asp Thr His Ala Leu Gln
530 535 540

Thr Pro His Leu Asp Thr Pro Gly Arg Arg Lys Arg Arg Ser Gly Asp
545 550 555 560

Leu Leu Pro Gln Gly Ala Thr Leu Leu Arg Trp Ile Ser Ala Gln Leu
565 570 575

His Ser Tyr Gln Phe Ile Pro Glu Leu Lys Glu Ala Ser Asp Val Phe
580 585 590

Arg Asn Gly Arg Val Leu Cys Ala Leu Ile Asn Arg Tyr Arg Pro Asp
595 600 605

Leu Ile Asp Tyr Ala Ala Thr Lys Asp Met Ser Pro Val Glu Cys Asn
610 615 620

Glu Leu Ser Phe Ala Val Leu Glu Arg Glu Leu His Ile Asp Arg Val
625 630 635 640

Met Ser Ala Lys Gln Ser Leu Asp Leu Thr Glu Leu Glu Ser Arg Ile
645 650 655

Trp Leu Asn Tyr Leu Asp Gln Ile Cys Asp Leu Phe Arg Gly Glu Ile
660 665 670

Pro His Ile Lys His Pro Lys Met Asp Phe Ser Asp Leu Arg Gln Lys
675 680 685

Tyr Arg Ile Asn His Thr His Ala Gln Pro Asp Phe Ser Lys Leu Leu
690 695 700

Ala Thr Lys Pro Lys Ala Lys Ser Pro Met Gln Asp Ala Val Asp Ile
705 710 715 720

Pro Thr Thr Val Gln Arg Arg Ser Val Leu Glu Glu Glu Arg Ala Lys
725 730 735

Arg Gln Arg Arg His Glu Gln Leu Leu Asn Ile Gly Gly Gly Ala Ala
740 745 750

Gly Ala Ala Ala Gly Val Ala Gly Ser Gly Thr Gly Thr Thr Thr Gln
755 760 765

Gly Gln Asn Asp Thr Pro Arg Arg Ser Lys Lys Arg Arg Gln Val Asp
770 775 780

Lys Thr Ala Asn Ile Glu Glu Arg Gln Gln Arg Leu Gln Glu Ile Glu
785 790 795 800

Glu Asn Arg Gln Glu Arg Met Ser Lys Arg Arg Gln Gln Arg Cys His
805 810 815

Gln Thr Gln Asn Phe Tyr Lys Ser Leu Gln Leu Leu Gln Ala Gly Lys
820 825 830

Leu Leu Arg Glu Gly Gly Glu Ala Gly Val Ala Glu Asp Gly Thr Pro
835 840 845

Phe Glu Asp Tyr Ser Ile Phe Leu Tyr Arg Gln Gln Ala Pro Val Phe
850 855 860

Asn Asp Arg Val Lys Asp Leu Glu Arg Lys Leu Leu Phe Pro Asp Arg
865 870 875 880

Glu Arg Gly Asp Ile Pro Ser Ala Leu Pro Arg Thr Ala Asp Glu Gln
885 890 895

Phe Ser Asp Arg Ile Lys Asn Met Glu Gln Arg Met Thr Gly Arg Gly
900 905 910

Gly Leu Gly Gly Asp Lys Lys Pro Lys Asp Leu Met Arg Ala Ile Gly
915 920 925

Lys Ile Asp Ser Asn Asp Trp Asn Val Arg Glu Ile Glu Lys Lys Ile
930 935 940

Glu Leu Ser Lys Lys Thr Glu Ile His Gly Pro Lys Gly Arg Glu Lys
945 950 955 960

Val Pro Lys Trp Ser Lys Glu Gln Phe Gln Ala Arg Gln His Lys Met
965 970 975

Ser Lys Pro Gln Arg Gln Asp Ser Arg Glu Ala Glu Lys Phe Lys Asp
980 985 990

Ile Asp Gln Thr Ile Arg Asn Leu Asp Lys Gln Leu Lys Glu Gly His
995 1000 1005

Asn Leu Asp Val Gly Glu Arg Gly Arg Asn Lys Val Ala Ser Ile
1010 1015 1020

Ala Gly Gln Phe Gly Lys Lys Asp Glu Ala Asn Ser Asp Glu Lys
1025 1030 1035

Asn Ala Gly Ser Ser Asn Ala Thr Thr Asn Thr Asn Asn Thr Val
1040 1045 1050

Ile Pro Lys Ser Ser Ser Lys Val Ala Leu Ala Phe Lys Lys Gln
1055 1060 1065

Ala Ala Ser Glu Lys Cys Arg Phe Cys Lys Gln Thr Val Tyr Leu
1070 1075 1080

Met Glu Lys Thr Thr Val Glu Gly Leu Val Leu His Arg Asn Cys

1085 1090 1095

Leu Lys Cys His His Cys His Thr Asn Leu Arg Leu Gly Gly Tyr
1100 1105 1110

Ala Phe Asp Arg Asp Asp Pro Gln Gly Arg Phe Tyr Cys Thr Gln
1115 1120 1125

His Phe Arg Leu Pro Pro Lys Pro Leu Pro Gln Arg Thr Asn Lys
1130 1135 1140

Ala Arg Lys Ser Ala Ala Ala Gln Pro Ala Ser Pro Ala Val Pro
1145 1150 1155

Pro Thr Ala Gly Ser Val Pro Thr Ala Ala Ala Thr Ser Glu His
1160 1165 1170

Met Asp Thr Thr Pro Pro Arg Asp Gln Val Asp Leu Leu Gln Thr
1175 1180 1185

Ser Arg Ala Asn Ala Ser Ala Asp Ala Met Ser Asp Asp Glu Ala
1190 1195 1200

Asn Val Ile Asp Glu His Glu Trp Ser Gly Arg Asn Phe Leu Pro
1205 1210 1215

Glu Ser Asn Asn Asp Ser Gln Ser Glu Leu Ser Ser Ser Asp Glu
1220 1225 1230

Ser Asp Thr Glu Ser Asp Ser Glu Met Phe Glu Glu Ala Asp Asp
1235 1240 1245

Ser Pro Phe Gly Ala Gln Thr Leu Gln Leu Ala Ser Asp Trp Ile
1250 1255 1260

Gly Lys Gln Tyr Cys Glu Asp Ser Asp Asp Ser Asp Asp Phe Tyr
1265 1270 1275

Asp Ser Ser Glu Gly Ile Ala Asp Asp Gly Lys Asp Asp Thr Glu
1280 1285 1290

Gly Glu Glu Phe Lys Lys Ala Arg Glu Leu Arg Arg Gln Glu Val
1295 1300 1305

Arg Leu Gln Pro Leu Pro Ala Asn Leu Pro Thr Asp Thr Glu Thr
1310 1315 1320

Glu Val Gln Thr Glu Ser Glu Ser Thr Ser Pro Asp Glu Val Glu
1325 1330 1335

Leu Asn Ser Ala Thr Glu Ile Ser Thr Asp Ser Glu Phe Asp Asn
1340 1345 1350

Asp Glu Ile Ile Arg Gln Ala Pro Lys Ile Phe Ile Asp Asp Thr
1355 1360 1365

His Leu Arg Lys Pro Thr Lys Val Gln Ile Lys Ser Thr Met Ile
1370 1375 1380

Gly Pro Asn Ala Ala Ser Ala Gly Leu His Gln Lys Gln Leu Ala
1385 1390 1395

Ala Arg Glu Lys Gly Gly Ser Tyr Leu Gln Lys Tyr Gln Pro Gln
1400 1405 1410

Pro Pro Leu Ser Gln Phe Lys Pro Leu Val Gln Val Asp Pro Thr
1415 1420 1425

Leu Leu Ile Gly Ser Gln Arg Ala Pro Leu Gln Asn Pro Arg Pro
1430 1435 1440

Gly Asp Tyr Leu Leu Asn Lys Thr Ala Ser Thr Glu Gly Ile Ala
1445 1450 1455

Ser Lys Lys Ser Leu Glu Leu Lys Lys Arg Tyr Leu Leu Gly Glu
1460 1465 1470

Pro Ala Asn Gly Asp Lys Ile Gln Lys Ser Gly Ser Thr Ser Val
1475 1480 1485

Leu Asp Ser Arg Ile Arg Ser Phe Gln Ser Asn Ile Ser Glu Cys
1490 1495 1500

Gln Lys Leu Leu Asn Pro Ser Ser Asp Ile Ser Ala Gly Met Arg
1505 1510 1515

Thr Phe Leu Asp Arg Thr Lys Leu Gly Glu Gly Ser Gln Thr Thr
1520 1525 1530

Pro Gly Gln Thr Asn Glu Leu Ile Arg Ser Ala Thr Ser Asn Val
1535 1540 1545

Ile Asn Asp Leu Arg Val Glu Leu Arg Ile Gln Lys Thr Asp Ser
1550 1555 1560

Ser His Ser Thr Asp Asn Glu Lys Glu Asn Val Phe Val Asn Cys
1565 1570 1575

Lys Asn Glu Leu Asn Lys Gly Met Glu Tyr Thr Asp Ala Val Asn
1580 1585 1590

Ala Thr Leu Leu Asp Gln Leu Ala Arg Lys Ser Ser Pro Thr Thr
1595 1600 1605

Pro Thr Asn Lys Thr Val Val Glu Val Ile Asp Leu Val Thr Pro
1610 1615 1620

Glu Lys Pro Ile Asp Ile Ile Asp Leu Thr Ala Leu Glu Thr Pro
1625 1630 1635

Lys Lys Gln Leu Val Asp Gly Ser Ala Met Asp Val Asp Glu Arg
1640 1645 1650

Leu Thr Pro Asp Ser Asn Lys Ile Ser Glu Leu Gln Gln Glu Val
1655 1660 1665

Lys Glu Glu Pro Lys Pro Asp Val Ser Arg Asp Val Lys Glu Cys
1670 1675 1680

Ile Pro Asp Ile Leu Gly His Ile Lys Glu Gly Thr Gly Ser Lys
1685 1690 1695

Glu Pro Gly Gly Glu Asp Gln Gln Ser Leu Leu Glu Gln Ser Asp
1700 1705 1710

Glu Glu Lys Arg Asp Ser Pro Glu Lys Asp Val Ala Glu His Glu
1715 1720 1725

Leu Tyr Glu Pro Asp Ser Val Gln Ile Gln Val Pro Asn Ile Pro
1730 1735 1740

Trp Glu Lys Ser Lys Pro Glu Val Met Ser Thr Thr Gly Ser Ser
1745 1750 1755

Gly Ser Ile Cys Ser Ser Ser Asp Ser Ser Ser Ile Glu Asp Ile
1760 1765 1770

Gln His Tyr Ile Leu Glu Ser Thr Thr Ser Pro Asp Thr Gln Thr
1775 1780 1785

Val Gly Gly Lys His Asn Val Pro Arg Leu Glu Val His Asp Thr
1790 1795 1800

Ser Gly Ala Leu Met Gln Val Asp Ser Leu Met Ile Val Asn Gly
1805 1810 1815

Lys Tyr Ile Gly Asp Pro Glu Asp Val Lys Phe Leu Asp Met Pro
1820 1825 1830

Ala Asn Val Ile Val Pro Pro Ala Pro Ala Leu Lys Thr Asn Glu
1835 1840 1845

Leu Asp Met Glu Asp Asp Gln Glu Ala Glu Ala Glu Pro Val Thr
1850 1855 1860

Ala Thr Pro Glu Pro Val Glu Cys Thr Val Ile Glu Ala Glu Arg
1865 1870 1875

Arg Val Thr Ala Pro Pro Pro Leu Pro Glu Met Gly Pro Pro Lys
1880 1885 1890

Leu Lys Phe Asp Ser Lys Asn Glu Asn Lys Ile Glu Ser Leu Lys
1895 1900 1905

Asn Leu Pro Leu Ile Val Glu Ser Asn Val Glu His Ser Gln Ala
1910 1915 1920

Val Lys Pro Ile Thr Leu Asn Leu Ser Asn Leu Ala Arg Thr Pro
1925 1930 1935

Asp Thr Pro Thr Thr Pro Thr Ala His Asp Ser Asp Lys Thr Pro
1940 1945 1950

Thr Gly Glu Ile Leu Ser Arg Gly Ser Asp Ser Glu Thr Glu His
1955 1960 1965

Thr Gly Thr Gly Gln Val Leu Thr Glu Thr Glu Leu Ser Asp Trp
1970 1975 1980

Thr Ala Asp Asp Cys Ile Ser Glu Asn Phe Val Asp Leu Glu Phe
1985 1990 1995

Ala Leu Asn Ser Asn Lys Gly Thr Ile Lys Arg Arg Lys Asp Arg

2000 2005 2010

Arg Arg Ser Gly Ala Ser Lys Leu Pro Ser Gly Asn Glu Val Ile
2015 2020 2025

His Glu Leu Ala Arg Gln Ala Pro Val Val Gln Met Asp Gly Ile
2030 2035 2040

Leu Ser Ala Ile Asp Ile Asp Asp Ile Glu Phe Met Asp Thr Gly
2045 2050 2055

Ser Glu Gly Ser Cys Ala Glu Ala Tyr Pro Ala Thr Asn Thr Ala
2060 2065 2070

Leu Ile Gln Asn Arg Gly Tyr Met Glu Tyr Ile Glu Ala Glu Pro
2075 2080 2085

Lys Lys Thr Thr Arg Lys Ala Ala Pro Pro Ser Ser Tyr Pro Gly
2090 2095 2100

Asn Leu Pro Pro Leu Met Thr Lys Arg Asp Glu Lys Leu Gly Val
2105 2110 2115

Asp Tyr Ile Glu Gln Gly Ala Tyr Ile Met His Asp Asp Ala Lys
2120 2125 2130

Thr Pro Val Asn Glu Val Ala Pro Ala Met Thr Gln Ser Leu Thr
2135 2140 2145

Asp Ser Ile Thr Leu Asn Glu Leu Asp Asp Asp Ser Met Ile Ile
2150 2155 2160

Ser Gln Thr Gln Pro Thr Thr Thr Glu Glu Ser Glu Ala Leu Thr
2165 2170 2175

Val Val Thr Ser Pro Leu Asp Thr Ser Ser Pro Arg Val Leu Asp
2180 2185 2190

Gln Phe Ala Ser Met Leu Ala Ala Gly Lys Gly Asp Ser Thr Pro
2195 2200 2205

Ser Ser Ser Glu Gln Gln Pro Lys Thr Ser Thr Val Thr Ser Ser
2210 2215 2220

Ser Thr Gly Pro Asn Ser Ser Thr Thr Gly Asn Val Ser Lys Glu
2225 2230 2235

Pro Gln Glu Glu Asp Leu Gln Ile Gln Phe Glu Tyr Val Arg Ala
2240 2245 2250

Leu Gln Gln Arg Ile Ser Gln Ile Ser Thr Gln Arg Arg Lys Ser
2255 2260 2265

Ser Lys Gly Glu Ala Pro Asn Leu Gln Leu Asn Ser Ser Ala Pro
2270 2275 2280

Val Ile Glu Ser Ala Glu Asp Pro Ala Lys Pro Ala Glu Glu Pro
2285 2290 2295

Leu Val Ser Met Arg Pro Arg Thr Thr Ser Ile Ser Gly Lys Val
2300 2305 2310

Pro Glu Ile Pro Thr Leu Ser Ser Lys Leu Glu Glu Ile Thr Lys
2315 2320 2325

Glu Arg Thr Lys Gln Lys Asp Leu Ile His Asp Leu Val Met Asp
2330 2335 2340

Lys Leu Gln Ser Lys Lys Gln Leu Asn Ala Glu Lys Arg Leu His
2345 2350 2355

Arg Ser Arg Gln Arg Ser Leu Leu Thr Ser Gly Tyr Ala Ser Gly
2360 2365 2370

Ser Ser Leu Ser Pro Thr Pro Lys Leu Ala Ala Ala Cys Ser Pro
2375 2380 2385

Gln Asp Ser Asn Cys Ser Ser Gln Ala His Tyr His Ala Ser Thr
2390 2395 2400

Ala Glu Glu Ala Pro Lys Pro Pro Ala Glu Arg Pro Leu Gln Lys
2405 2410 2415

Ser Ala Thr Ser Thr Tyr Val Ser Pro Tyr Arg Thr Val Gln Ala
2420 2425 2430

Pro Thr Arg Ser Ala Asp Leu Tyr Lys Pro Arg Pro Phe Ser Glu
2435 2440 2445

His Ile Asp Ser Asn Ala Leu Ala Gly Tyr Lys Leu Gly Lys Thr
2450 2455 2460

Ala Ser Phe Asn Gly Gly Lys Leu Gly Asp Phe Ala Lys Pro Ile
2465 2470 2475

Ala Pro Ala Arg Val Asn Arg Gly Gly Gly Val Ala Thr Ala Asp
2480 2485 2490

Ile Ala Asn Ile Ser Ala Ser Thr Glu Asn Leu Arg Ser Glu Ala
2495 2500 2505

Arg Ala Arg Ala Arg Leu Lys Ser Asn Thr Glu Leu Gly Leu Ser
2510 2515 2520

Pro Glu Glu Lys Met Gln Leu Ile Arg Ser Arg Leu His Tyr Asp
2525 2530 2535

Gln Asn Arg Ser Leu Lys Pro Lys Gln Leu Glu Glu Met Pro Ser
2540 2545 2550

Gly Asp Leu Ala Ala Arg Ala Arg Lys Met Ser Ala Ser Lys Ser
2555 2560 2565

Val Asn Asp Leu Ala Tyr Met Val Gly Gln Gln Gln Gln Gln Gln
2570 2575 2580

Val Glu Lys Asp Ala Val Leu Gln Ala Lys Ala Ala Asp Phe Thr
2585 2590 2595

Ser Asp Pro Asn Leu Ala Ser Gly Gly Gln Glu Lys Ala Gly Lys
2600 2605 2610

Thr Lys Ser Gly Arg Arg Pro Lys Asp Pro Glu Arg Arg Lys Ser
2615 2620 2625

Leu Ile Gln Ser Leu Ser Ser Phe Phe Gln Lys Gly Ser Gly Ser
2630 2635 2640

Ala Ala Ser Ser Ser Lys Glu Gln Gly Gly Ala Val Ala Ala Val
2645 2650 2655

His Ser Glu Gln Ser Glu Arg Pro Gly Thr Ser Ser Ser Gly Thr
2660 2665 2670

Pro Thr Ile Ser Asp Ala Ala Gly Gly Gly Gly Gly Gly Gly Gly
2675 2680 2685

Val Phe Ser Arg Phe Arg Ile Ser Pro Lys Ser Lys Glu Lys Ser
2690 2695 2700

Lys Ser Cys Phe Asp Leu Arg Asn Phe Gly Phe Gly Asp Lys Asp
2705 2710 2715

Met Leu Val Cys Asn Ala Ala Ser Pro Ala Gly Ala Thr Ser Ala
2720 2725 2730

Ser Gln Lys Asn His Ser Gln Glu Tyr Leu Asn Thr Thr Asn Asn
2735 2740 2745

Ser Arg Tyr Arg Lys Gln Thr Asn Thr Ala Lys Pro Lys Pro Glu
2750 2755 2760

Ser Phe Ser Ser Ser Ser Pro Gln Leu Tyr Ile His Lys Pro His
2765 2770 2775

His Leu Ala Ala Ala His Pro Ser Ala Leu Asp Asp Gln Thr Pro
2780 2785 2790

Pro Pro Ile Pro Pro Leu Pro Leu Asn Tyr Gln Arg Ser Asp Asp
2795 2800 2805

Glu Ser Tyr Ala Asn Glu Thr Arg Glu His Lys Lys Gln Arg Ala
2810 2815 2820

Ile Ser Lys Ala Ser Arg Gln Ala Glu Leu Lys Arg Leu Arg Ile
2825 2830 2835

Ala Gln Glu Ile Gln Arg Glu Gln Glu Glu Ile Glu Val Gln Leu
2840 2845 2850

Lys Asp Leu Glu Ala Arg Gly Val Leu Ile Glu Lys Ala Leu Arg
2855 2860 2865

Gly Glu Ala Gln Asn Ile Glu Asn Leu Asp Ala Thr Lys Asp Asn
2870 2875 2880

Asp Glu Lys Leu Leu Lys Glu Leu Leu Glu Ile Trp Arg Asn Ile
2885 2890 2895

Thr Ala Leu Lys Lys Arg Asp Glu Glu Leu Thr Ile Arg Gln Gln
2900 2905 2910

Glu Leu Gln Leu Glu Tyr Arg His Ala Gln Leu Lys Glu Glu Leu

2915 2920 2925

Asn Leu Arg Leu Ser Cys Asn Lys Leu Asp Lys Ser Ser Ala Asp
2930 2935 2940

Val Ala Ala Glu Gly Ala Ile Leu Asn Glu Met Leu Glu Ile Val
2945 2950 2955

Ala Lys Arg Ala Ala Leu Arg Pro Thr Ala Ser Gln Leu Asp Leu
2960 2965 2970

Thr Ala Ala Gly Ser Ala Ser Thr Ser Ala Glu Ala Thr Gly Ile
2975 2980 2985

Lys Leu Thr Gly Gln Pro His Asp His Glu Glu Ser Ile Ile
2990 2995 3000

<210> 21
<211> 1048
<212> PRT
<213> Mouse

<400> 21

Met Ala Ser Pro Ala Ser Thr Asn Pro Ala His Asp His Phe Glu Thr
1 5 10 15

Phe Val Gln Ala Gln Leu Cys Gln Asp Val Leu Ser Ser Phe Gln Gly
20 25 30

Leu Cys Arg Ala Leu Gly Val Glu Ser Gly Gly Gly Leu Ser Gln Tyr
35 40 45

His Lys Ile Lys Ala Gln Leu Asn Tyr Trp Ser Ala Lys Ser Leu Trp
50 55 60

Ala Lys Leu Asp Lys Arg Ala Ser Gln Pro Val Tyr Gln Gln Gly Gln
65 70 75 80

Ala Cys Thr Asn Thr Lys Cys Leu Val Val Gly Ala Gly Pro Cys Gly
85 90 95

Leu Arg Ala Ala Val Glu Leu Ala Leu Leu Gly Ala Arg Val Val Leu
100 105 110

Val Glu Lys Arg Ile Lys Phe Ser Arg His Asn Val Leu His Leu Trp
115 120 125

Pro Phe Thr Ile His Asp Leu Arg Ala Leu Gly Ala Lys Lys Phe Tyr
130 135 140

Gly Arg Phe Cys Thr Gly Thr Leu Asp His Ile Ser Ile Arg Gln Leu
145 150 155 160

Gln Leu Leu Leu Leu Lys Val Ala Leu Leu Leu Gly Val Glu Ile His
165 170 175

Trp Gly Val Lys Phe Thr Gly Leu Gln Pro Pro Pro Arg Lys Gly Ser
180 185 190

Gly Trp Arg Ala Gln Leu Gln Pro Asn Pro Pro Ala Gln Leu Ala Ser
195 200 205

Tyr Glu Phe Asp Val Leu Ile Ser Ala Ala Gly Gly Lys Phe Val Pro
210 215 220

Glu Gly Phe Thr Ile Arg Glu Met Arg Gly Lys Leu Ala Ile Gly Ile
225 230 235 240

Thr Ala Asn Phe Val Asn Gly Arg Thr Val Glu Glu Thr Gln Val Pro
245 250 255

Glu Ile Ser Gly Val Ala Arg Ile Tyr Asn Gln Lys Phe Phe Gln Ser
260 265 270

Leu Leu Lys Ala Thr Gly Ile Asp Leu Glu Asn Ile Val Tyr Tyr Lys
275 280 285

Asp Glu Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln Cys Leu Leu
290 295 300

Arg Leu Gly Val Leu Arg Gln Asp Leu Ser Glu Thr Asp Gln Leu Leu
305 310 315 320

Gly Lys Ala Asn Val Val Pro Glu Ala Leu Gln Arg Phe Ala Arg Ala
325 330 335

Ala Ala Asp Phe Ala Thr His Gly Lys Leu Gly Lys Leu Glu Phe Ala
340 345 350

Gln Asp Ala Arg Gly Arg Pro Asp Val Ala Ala Phe Asp Phe Thr Ser
355 360 365

Met Met Arg Ala Glu Ser Ser Ala Arg Val Gln Glu Lys His Gly Ala
370 375 380

Arg Leu Leu Leu Gly Leu Val Gly Asp Cys Leu Val Glu Pro Phe Trp
385 390 395 400

Pro Leu Gly Thr Gly Val Ala Arg Gly Phe Leu Ala Ala Phe Asp Ala
405 410 415

Ala Trp Met Val Lys Arg Trp Ala Glu Gly Ala Gly Pro Leu Glu Val
420 425 430

Leu Ala Glu Arg Glu Ser Leu Tyr Gln Leu Leu Ser Gln Thr Ser Pro
435 440 445

Glu Asn Met His Arg Asn Val Ala Gln Tyr Gly Leu Asp Pro Ala Thr
450 455 460

Arg Tyr Pro Asn Leu Asn Leu Arg Ala Val Thr Pro Asn Gln Val Gln
465 470 475 480

Asp Leu Tyr Asp Met Met Asp Lys Glu His Ala Gln Arg Lys Ser Asp
485 490 495

Glu Pro Asp Ser Arg Lys Thr Thr Thr Gly Ser Ala Gly Thr Glu Glu
500 505 510

Leu Leu His Trp Cys Gln Glu Gln Thr Ala Gly Phe Pro Gly Val His
515 520 525

Val Thr Asp Phe Ser Ser Ser Trp Ala Asp Gly Leu Ala Leu Cys Ala
530 535 540

Leu Val His His Leu Gln Pro Gly Leu Leu Glu Pro Ser Glu Leu Gln
545 550 555 560

Gly Met Gly Ala Leu Glu Ala Thr Thr Trp Ala Leu Arg Val Ala Glu
565 570 575

His Glu Leu Gly Ile Thr Pro Val Leu Ser Ala Gln Ala Val Met Ala
580 585 590

Gly Ser Asp Pro Leu Gly Leu Ile Ala Tyr Leu Ser His Phe His Ser
595 600 605

Ala Phe Lys Asn Thr Ser His Ser Ser Gly Leu Val Ser Gln Pro Ser

610 615 620

Gly Thr Pro Ser Ala Ile Leu Phe Leu Gly Lys Leu Gln Arg Ser Leu
625 630 635 640

Gln Arg Thr Arg Ala Lys Val Asp Glu Glu Thr Pro Ser Thr Glu Glu
645 650 655

Pro Pro Val Ser Glu Pro Ser Met Ser Pro Asn Thr Pro Glu Leu Ser
660 665 670

Glu His Gln Glu Ala Gly Ala Glu Glu Leu Cys Glu Leu Cys Gly Lys
675 680 685

His Leu Tyr Ile Leu Glu Arg Phe Cys Val Asp Gly His Phe Phe His
690 695 700

Arg Ser Cys Phe Cys Cys His Thr Cys Glu Ala Thr Leu Trp Pro Gly
705 710 715 720

Gly Tyr Gly Gln His Pro Gly Asp Gly His Phe Tyr Cys Leu Gln His
725 730 735

Leu Pro Gln Glu Asp Gln Lys Glu Ala Asp Asn Asn Gly Ser Leu Glu
740 745 750

Ser Gln Glu Leu Pro Thr Pro Gly Asp Ser Asn Met Gln Pro Asp Pro
755 760 765

Ser Ser Pro Pro Val Thr Arg Val Ser Pro Val Pro Ser Pro Ser Gln
770 775 780

Pro Ala Arg Arg Leu Ile Arg Leu Ser Ser Leu Glu Arg Leu Arg Leu
785 790 795 800

Ser Ser Leu Asn Ile Ile Pro Asp Ser Gly Ala Glu Pro Pro Pro Lys
805 810 815

Pro Pro Arg Ser Cys Ser Asp Leu Ala Arg Glu Ser Leu Lys Ser Ser
820 825 830

Phe Val Gly Trp Gly Val Pro Val Gln Ala Pro Gln Val Pro Glu Ala
835 840 845

Ile Glu Lys Gly Asp Asp Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
850 855 860

Glu Glu Glu Pro Leu Pro Pro Leu Glu Pro Glu Leu Glu Gln Thr Leu
865 870 875 880

Leu Thr Leu Ala Lys Asn Pro Gly Ala Met Thr Lys Tyr Pro Thr Trp
885 890 895

Arg Arg Thr Leu Met Arg Arg Ala Lys Glu Glu Glu Met Lys Arg Phe
900 905 910

Cys Lys Ala Gln Ala Ile Gln Arg Arg Leu Asn Glu Ile Glu Ala Thr
915 920 925

Met Arg Glu Leu Glu Ala Glu Gly Thr Lys Leu Glu Leu Ala Leu Arg
930 935 940

Lys Glu Ser Ser Ser Pro Glu Gln Gln Lys Lys Leu Trp Leu Asp Gln
945 950 955 960

Leu Leu Arg Leu Ile Gln Lys Lys Asn Ser Leu Val Thr Glu Glu Ala
965 970 975

Glu Leu Met Ile Thr Val Gln Glu Leu Asp Leu Glu Glu Lys Gln Arg
980 985 990

Gln Leu Asp His Glu Leu Arg Gly Tyr Met Asn Arg Glu Glu Thr Met
995 1000 1005

Lys Thr Glu Ala Asp Leu Gln Ser Glu Asn Gln Val Leu Arg Lys
1010 1015 1020

Leu Leu Glu Val Val Asn Gln Arg Asp Ala Leu Ile Gln Phe Gln
1025 1030 1035

Glu Glu Arg Arg Leu Arg Glu Met Pro Ala
1040 1045

<210> 22
<211> 1480
<212> PRT
<213> Mouse

<400> 22

Met Gly Glu Asn Glu Asp Glu Lys Gln Ala Gln Ala Ser Gln Val Phe
1 5 10 15

```
Glu Asn Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Gln Ala Phe
            20                  25                  30

Asn Ile Leu Thr Cys Leu Leu Asp Leu Asp Pro Leu Asp His Arg Asn
        35                  40                  45

Phe Tyr Ser Gln Leu Lys Ser Lys Val Asn Thr Trp Lys Ala Lys Ala
    50                  55                  60

Leu Trp His Lys Leu Asp Lys Arg Gly Ser His Lys Glu Tyr Lys Arg
65                  70                  75                  80

Gly Lys Ala Cys Ser Asn Thr Lys Cys Leu Ile Val Gly Gly Gly Pro
                85                  90                  95

Cys Gly Leu Arg Thr Ala Ile Glu Leu Ala Tyr Leu Gly Ala Lys Val
            100                 105                 110

Val Val Val Glu Lys Arg Asp Thr Phe Ser Arg Asn Asn Val Leu His
        115                 120                 125

Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala Lys Lys
    130                 135                 140

Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His Ile Ser Ile Arg
145                 150                 155                 160

Gln Leu Gln Leu Ile Leu Phe Lys Val Ala Leu Met Leu Gly Val Glu
                165                 170                 175

Val His Val Asn Val Glu Phe Val Arg Val Leu Glu Pro Pro Glu Asp
            180                 185                 190

Gln Glu Asn Gln Lys Val Gly Trp Arg Ala Glu Phe Leu Pro Ala Asp
        195                 200                 205

His Ala Leu Ser Asp Phe Glu Phe Asp Val Ile Ile Gly Ala Asp Gly
    210                 215                 220

His Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg Gly Lys
225                 230                 235                 240

Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Ser Thr Ala
                245                 250                 255

Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln
```

260 265 270

Lys Phe Phe Gln Asp Leu Lys Glu Glu Thr Gly Ile Asp Leu Glu Asn
275 280 285

Ile Val Tyr Tyr Lys Asp Ser Thr His Tyr Phe Val Met Thr Ala Lys
290 295 300

Lys Gln Ser Leu Leu Asp Lys Gly Val Ile Leu Asn Asp Tyr Ile Asp
305 310 315 320

Thr Glu Met Leu Leu Cys Ser Glu Asn Val Asn Gln Asp Asn Leu Leu
325 330 335

Ser Tyr Ala Arg Glu Ala Ala Asp Phe Ala Thr Asn Tyr Gln Leu Pro
340 345 350

Ser Leu Asp Phe Ala Ile Asn His Asn Gly Gln Pro Asp Val Ala Met
355 360 365

Phe Asp Phe Thr Ser Met Tyr Ala Ser Glu Asn Ala Ala Leu Met Arg
370 375 380

Glu Arg Gln Ala His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu
385 390 395 400

Leu Glu Pro Phe Trp Pro Met Gly Thr Gly Cys Ala Arg Gly Phe Leu
405 410 415

Ala Ala Phe Asp Thr Ala Trp Met Val Lys Ser Trp Asp Gln Gly Thr
420 425 430

Pro Pro Leu Glu Val Leu Ala Glu Arg Glu Ser Leu Tyr Arg Leu Leu
435 440 445

Pro Gln Thr Thr Pro Glu Asn Ile Asn Lys Asn Phe Glu Gln Tyr Thr
450 455 460

Leu Asp Pro Ala Thr Arg Tyr Pro Asn Leu Asn Leu His Cys Val Arg
465 470 475 480

Pro His Gln Val Lys His Leu Tyr Ile Thr Lys Glu Met Asp Arg Phe
485 490 495

Pro Leu Glu Arg Trp Gly Ser Val Arg Arg Ser Val Ser Leu Ser Arg
500 505 510

Arg Glu Ser Asp Ile Arg Pro Asn Lys Leu Leu Thr Trp Cys Gln Gln
515 520 525

Gln Thr Lys Gly Tyr Gln His Val Arg Val Thr Asp Leu Thr Thr Ser
530 535 540

Trp Arg Ser Gly Leu Ala Leu Cys Ala Ile Ile His Ser Phe Arg Pro
545 550 555 560

Glu Leu Ile Asn Phe Asp Ser Leu Asn Glu Asp Asp Ala Val Glu Asn
565 570 575

Asn Gln Leu Ala Phe Asp Val Ala Lys Arg Glu Phe Gly Ile Leu Pro
580 585 590

Val Thr Thr Gly Lys Glu Met Ala Ser Thr Gln Glu Pro Asp Lys Leu
595 600 605

Ser Met Val Met Tyr Leu Ser Lys Phe Tyr Glu Leu Phe Arg Gly Thr
610 615 620

Pro Leu Arg Pro Met Asp Ser Trp Arg Lys Asn Tyr Gly Glu Asn Ala
625 630 635 640

Asp Phe Gly Leu Gly Lys Thr Phe Ile Gln Asn Asn Tyr Leu Asn Leu
645 650 655

Thr Leu Pro Arg Lys Arg Thr Pro Arg Val Asp Thr Gln Thr Glu Glu
660 665 670

Asn Asp Met Asn Lys Arg Arg Arg Gln Gly Phe Asn His Leu Glu Glu
675 680 685

Leu Pro Ser Phe Ser Ser Arg Ser Leu Gly Ser Ser Gln Glu Tyr Ala
690 695 700

Lys Glu Ser Gly Ser Gln Asn Lys Val Lys His Met Ala Asn Gln Leu
705 710 715 720

Leu Ala Lys Phe Glu Glu Asn Thr Arg Asn Pro Ser Val Val Lys Gln
725 730 735

Glu Ser Pro Arg Lys Ala Phe Pro Leu Ser Leu Gly Gly Arg Asp Thr
740 745 750

Cys Tyr Phe Cys Lys Lys Arg Val Tyr Met Ile Glu Arg Leu Ser Ala
755 760 765

Glu Gly His Phe Phe His Gln Glu Cys Phe Arg Cys Ser Val Cys Ser
770 775 780

Ala Thr Leu Arg Leu Ala Ala Tyr Ala Phe Asp Cys Asp Glu Gly Lys
785 790 795 800

Phe Tyr Cys Lys Pro His Phe Val His Cys Lys Thr Ser Ser Lys Gln
805 810 815

Arg Lys Arg Arg Ala Glu Leu Asn Gln Gln Arg Glu Glu Glu Gly Thr
820 825 830

Trp Gln Glu Gln Glu Ala Pro Arg Arg Asp Val Pro Thr Glu Ser Ser
835 840 845

Cys Ala Val Ala Ala Ile Ser Thr Pro Glu Gly Ser Pro Pro Gly Thr
850 855 860

Ser Thr Ser Phe Phe Arg Lys Ala Leu Ser Trp Pro Leu Arg Leu Thr
865 870 875 880

Arg Gly Leu Leu Asn Leu Pro Gln Ser Leu Leu Arg Trp Met Gln Gly
885 890 895

Leu Gln Glu Ala Ala Gly His His Val Arg Asp Asn Ala His Asn Tyr
900 905 910

Cys Phe Met Phe Glu Leu Leu Ser Leu Gly Leu Leu Leu Leu Trp Ala
915 920 925

Phe Ser Lys Val Leu Ala Ala Met Tyr Arg Glu Ser Glu Glu Ser Leu
930 935 940

Glu Asn Ile Arg Ser Trp Leu Leu Arg Phe Ile Pro Val Lys Leu Gln
945 950 955 960

Met Gly Gln Pro Gly Gly Pro Glu Leu Ser Lys Glu Arg Lys Leu Gly
965 970 975

Leu Lys Lys Leu Val Leu Thr Glu Glu Gln Lys Asn Lys Leu Leu Asp
980 985 990

Trp Ser Asp Cys Thr Gln Glu His Lys Thr Gly Glu Gln Leu Ser Gln
995 1000 1005

Glu Ser Ala Glu Asn Ile Arg Gly Gly Ser Leu Lys Pro Thr Cys
1010 1015 1020

Ser Ser Thr Leu Ser Gln Ala Val Lys Glu Lys Leu Leu Ser Gln
1025 1030 1035

Lys Lys Ala Leu Gly Gly Met Arg Thr Pro Ala Val Lys Ala Pro
1040 1045 1050

Gln Glu Arg Glu Val Pro Pro Pro Lys Ser Pro Leu Lys Leu Ile
1055 1060 1065

Ala Asn Ala Ile Leu Arg Ser Leu Leu His Asn Ser Glu Ala Gly
1070 1075 1080

Lys Lys Thr Ser Pro Lys Pro Glu Ser Lys Thr Leu Pro Arg Gly
1085 1090 1095

Gln Pro His Ala Arg Ser Phe Ser Leu Arg Lys Leu Gly Ser Ser
1100 1105 1110

Lys Asp Gly Asp Gln Gln Ser Pro Gly Arg His Met Ala Lys Lys
1115 1120 1125

Ala Ser Ala Phe Phe Ser Leu Ala Ser Pro Thr Ser Lys Val Ala
1130 1135 1140

Gln Ala Ser Asp Leu Ser Leu Pro Asn Ser Ile Leu Arg Ser Arg
1145 1150 1155

Ser Leu Pro Ser Arg Pro Ser Lys Met Phe Phe Ser Thr Thr Pro
1160 1165 1170

His Ser Lys Val Glu Asp Val Pro Thr Leu Leu Glu Lys Val Ser
1175 1180 1185

Leu Gln Asp Ala Thr His Ser Pro Lys Thr Gly Ala Ser His Ile
1190 1195 1200

Ser Ser Leu Gly Leu Lys Asp Lys Ser Phe Glu Ser Phe Leu Gln
1205 1210 1215

Glu Cys Lys Gln Arg Lys Asp Ile Gly Asp Phe Phe Asn Ser Pro

1220 1225 1230

Lys Glu Glu Gly Pro Pro Gly Asn Arg Val Pro Ser Leu Glu Lys
1235 1240 1245

Leu Val Gln Pro Val Gly Ser Thr Ser Met Gly Gln Val Ala His
1250 1255 1260

Pro Ser Ser Thr Gly Gln Asp Ala His Pro Val Ala Pro Val Thr
1265 1270 1275

Glu Ala Thr Ser Ser Pro Thr Ser Ser Ser Ala Glu Glu Glu Ala
1280 1285 1290

Asp Ser Gln Leu Ser Leu Arg Ile Lys Glu Lys Ile Leu Arg Arg
1295 1300 1305

Arg Arg Lys Leu Glu Lys Gln Ser Ala Lys Gln Glu Glu Leu Lys
1310 1315 1320

Arg Leu His Lys Ala Gln Ala Ile Gln Arg Gln Leu Glu Glu Val
1325 1330 1335

Glu Glu Arg Gln Arg Thr Leu Ala Ile Gln Gly Val Lys Leu Glu
1340 1345 1350

Lys Val Leu Arg Gly Glu Ala Ala Asp Ser Gly Thr Gln Asp Glu
1355 1360 1365

Ala Gln Leu Leu Gln Glu Trp Phe Lys Leu Val Leu Glu Lys Asn
1370 1375 1380

Lys Leu Met Arg Tyr Glu Ser Glu Leu Leu Ile Met Ala Gln Glu
1385 1390 1395

Leu Glu Leu Glu Asp His Gln Ser Arg Leu Glu Gln Lys Leu Arg
1400 1405 1410

Gln Lys Met Leu Lys Asp Glu Gly Gln Lys Asp Glu Asn Asp Leu
1415 1420 1425

Lys Glu Glu Gln Glu Ile Phe Glu Glu Met Met Gln Val Ile Glu
1430 1435 1440

Gln Arg Asn Lys Leu Val Asp Ser Leu Glu Glu Gln Arg Val Lys
1445 1450 1455

Glu Arg Thr Gln Asp Gln His Phe Glu Asn Phe Val Leu Ser Arg
1460 1465 1470

Gly Cys Gln Leu Ser Arg Thr
1475 1480

<210> 23
<211> 1026
<212> PRT
<213> Mouse

<220>
<221> MISC_FEATURE
<222> (1016)..(1016)
<223> Xaa is any amino acid

<400> 23

Met Glu Glu Arg Lys Gln Glu Thr Thr Asn Gln Ala His Val Leu Phe
1 5 10 15

Asp Arg Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Arg Ala Phe
20 25 30

Gln Glu Leu Cys Asp His Leu Glu Leu Lys Pro Lys Asp Tyr Arg Ser
35 40 45

Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala Lys Ala
50 55 60

Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr Lys Lys
65 70 75 80

Gly Lys Ala Cys Thr Asn Thr Lys Cys Leu Ile Ile Gly Ala Gly Pro
85 90 95

Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Leu Leu Gly Ala Lys Val
100 105 110

Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val Leu His
115 120 125

Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala Lys Lys
130 135 140

Phe Tyr Gly Lys Phe Cys Ala Gly Ala Ile Asp His Ile Ser Ile Arg
145 150 155 160

Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly Ile Glu
165 170 175

Ile His Val Asn Val Glu Phe Gln Gly Leu Val Gln Pro Pro Glu Asp
180 185 190

Gln Glu Asn Glu Arg Ile Gly Trp Arg Ala Leu Val His Pro Lys Thr
195 200 205

His Pro Val Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly Asp Gly
210 215 220

Arg Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg Gly Lys
225 230 235 240

Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Thr Thr Ala
245 250 255

Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln
260 265 270

Lys Phe Phe Gln Glu Leu Arg Glu Thr Thr Gly Ile Asp Leu Glu Asn
275 280 285

Ile Val Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys
290 295 300

Lys Gln Ser Leu Leu Asp Lys Gly Val Ile Leu His Asp Tyr Thr Asp
305 310 315 320

Thr Glu Leu Leu Leu Ser Arg Glu Asn Val Asp Gln Glu Ala Leu Leu
325 330 335

Asn Tyr Ala Arg Glu Ala Ala Asp Phe Ser Thr Gln Gln Gln Leu Pro
340 345 350

Ser Leu Asp Phe Ala Ile Asn His Tyr Gly Gln Pro Asp Val Ala Met
355 360 365

Phe Asp Phe Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg
370 375 380

Glu Gln Asn Gly His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu
385 390 395 400

```
Leu Glu Pro Phe Trp Pro Met Gly Thr Gly Ile Ala Arg Gly Phe Leu
                405                 410                 415

Ala Ala Met Asp Ser Ala Trp Met Val Arg Ser Trp Ser Leu Gly Thr
            420                 425                 430

Ser Pro Leu Glu Val Leu Ala Glu Arg Glu Ser Ile Tyr Arg Leu Leu
        435                 440                 445

Pro Gln Thr Thr Pro Glu Asn Val Ser Lys Asn Phe Ser Gln Tyr Ser
    450                 455                 460

Ile Asp Pro Val Thr Arg Tyr Pro Asn Ile Asn Ile Asn Phe Leu Arg
465                 470                 475                 480

Pro Ser Gln Val Arg His Leu Tyr Asp Ser Gly Glu Thr Lys Asp Ile
                485                 490                 495

His Leu Glu Met Glu Asn Met Val Asn Pro Arg Thr Thr Pro Lys Leu
            500                 505                 510

Thr Arg Asn Glu Ser Val Ala Arg Ser Ser Lys Leu Leu Gly Trp Cys
        515                 520                 525

Gln Arg Gln Thr Glu Gly Tyr Ser Gly Val Asn Val Thr Asp Leu Thr
    530                 535                 540

Met Ser Trp Lys Ser Gly Leu Ala Leu Cys Ala Ile Ile His Arg Tyr
545                 550                 555                 560

Arg Pro Asp Leu Ile Asp Phe Asp Ser Leu Asp Glu Gln Asn Val Glu
                565                 570                 575

Lys Asn Asn Gln Leu Ala Phe Asp Ile Ala Glu Lys Glu Leu Gly Ile
            580                 585                 590

Ser Pro Ile Met Thr Gly Lys Glu Met Ala Ser Val Gly Glu Pro Asp
        595                 600                 605

Lys Leu Ser Met Val Met Tyr Leu Thr Gln Phe Tyr Glu Met Phe Lys
    610                 615                 620

Asp Ser Leu Ser Ser Ser Asp Thr Leu Asp Leu Asn Ala Glu Glu Lys
625                 630                 635                 640
```

Ala Val Leu Ile Ala Ser Thr Lys Ser Pro Ile Ser Phe Leu Ser Lys
645 650 655

Leu Gly Gln Thr Ile Ser Arg Lys Arg Ser Pro Lys Asp Lys Lys Glu
660 665 670

Lys Asp Ser Asp Gly Ala Gly Lys Arg Arg Lys Thr Ser Gln Ser Glu
675 680 685

Glu Glu Glu Pro Pro Arg Ser Tyr Lys Gly Glu Arg Pro Thr Leu Val
690 695 700

Ser Thr Leu Thr Asp Arg Arg Met Asp Ala Ala Val Gly Asn Gln Asn
705 710 715 720

Lys Val Lys Tyr Met Ala Thr Gln Leu Leu Ala Lys Phe Glu Glu Asn
725 730 735

Ala Pro Ala Gln Ser Thr Gly Val Arg Arg Gln Gly Ser Ile Lys Lys
740 745 750

Glu Phe Pro Gln Asn Leu Gly Gly Ser Asp Thr Cys Tyr Phe Cys Gln
755 760 765

Lys Arg Val Tyr Val Met Glu Arg Leu Ser Ala Glu Gly Lys Phe Phe
770 775 780

His Arg Ser Cys Phe Lys Cys Glu Tyr Cys Ala Thr Thr Leu Arg Leu
785 790 795 800

Ser Ala Tyr Ala Tyr Asp Ile Glu Asp Glu Phe Ser Pro Asn Phe Trp
805 810 815

Cys Ser Ala His Tyr His Val Pro Val Ala Leu Pro Ala Thr Val Met
820 825 830

Pro Met Cys Leu Leu Tyr His Pro Ser Gln Val Leu Val Cys Leu Glu
835 840 845

Gly Gly Pro Ala Phe Met Ser Pro Val Leu Phe Asn Asp Thr Asn Ser
850 855 860

Arg Gln Ala Lys Gln Glu Glu Leu Lys Arg Leu His Arg Ala Gln Ile
865 870 875 880

Ile Gln Arg Gln Leu Glu Gln Val Glu Glu Lys Gln Arg Gln Leu Glu

885 890 895

Glu Arg Gly Val Ala Val Glu Lys Ala Leu Arg Gly Glu Ala Gly Met
900 905 910

Gly Lys Lys Asp Asp Pro Lys Leu Met Gln Glu Trp Phe Lys Leu Val
915 920 925

Gln Glu Lys Asn Ala Met Val Arg Tyr Glu Ser Glu Leu Met Ile Phe
930 935 940

Ala Arg Glu Leu Glu Leu Glu Asp Arg Gln Ser Arg Leu Gln Gln Glu
945 950 955 960

Leu Arg Glu Arg Met Ala Val Glu Asp His Leu Lys Thr Glu Gly Glu
965 970 975

Leu Ser Glu Glu Lys Lys Ile Leu Asn Glu Met Leu Glu Val Val Glu
980 985 990

Gln Arg Asp Ser Leu Val Ala Leu Leu Glu Glu Gln Arg Leu Arg Glu
995 1000 1005

Lys Glu Glu Asp Lys Asp Leu Xaa Ala Ala Met Leu Cys Lys Gly
1010 1015 1020

Phe Ser Leu
1025

<210> 24
<211> 476
<212> PRT
<213> Anopheles gambiae

<400> 24

Glu Met Phe Leu His Phe Cys Ala Ala Thr Thr Met Lys Gln Ile Arg
1 5 10 15

Gly Leu Tyr Trp Asn Met Leu Asp Thr Ile Gly Leu Arg Pro Gly Pro
20 25 30

Leu Glu Glu Phe Tyr Pro Lys Met Lys Ala Ala Ile Arg Asp Trp Arg
35 40 45

Ala Gln Ala Leu Phe Lys Lys Phe Asp Ala Arg Ala Ala His Lys Val
50 55 60

Tyr Cys Lys Gly Arg Ala Ala Ser Lys Thr Arg Val Leu Ile Val Gly
65 70 75 80

Ala Gly Pro Cys Gly Leu Arg Thr Ala Ile Asp Ala Gln Leu Leu Gly
85 90 95

Ala Lys Val Val Val Val Val Glu Lys Arg Asp Arg Ile Ser Arg Asn
100 105 110

Asn Val Leu His Leu Trp Pro Phe Ile Ile His Asp Leu Lys Ala Leu
115 120 125

Gly Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His
130 135 140

Ile Ser Ile Arg Gln Leu Gln Cys Ile Leu Leu Lys Val Ala Leu Leu
145 150 155 160

Leu Gly Val Glu Met His Glu Gly Val Ser Phe Val Lys Glu Ile Glu
165 170 175

Pro Gly Asp Gly Tyr Gly Trp Arg Ala Ser Val Ser Pro Glu Asp His
180 185 190

Ala Val Ser His Tyr Glu Phe Asp Val Leu Ile Gly Ala Asp Gly Lys
195 200 205

Arg Asn Thr Leu Glu Gly Phe Gln Arg Lys Glu Phe Arg Gly Lys Leu
210 215 220

Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Lys Arg Thr Glu Ala Glu
225 230 235 240

Ala Met Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asp Gln Pro
245 250 255

Phe Phe Lys Ala Leu Tyr Glu Lys Thr Gly Cys Asp Leu Glu Asn Ile
260 265 270

Val Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys
275 280 285

His Ser Leu Leu His Arg Gly Val Ile Ile Lys Asp Leu Ser Asp Pro
290 295 300

Ala Glu Leu Leu Ala Pro Ser Asn Val Asp Lys Pro Lys Leu Tyr Glu
305 310 315 320

Tyr Ala Arg Asp Ala Ala Asn Phe Ala Thr Lys Tyr Gln Met Pro Asn
325 330 335

Leu Glu Phe Ala Val Asn His Tyr Gly Thr Pro Asp Val Ala Val Phe
340 345 350

Asp Phe Thr Ser Ile Phe Ala Ala His Asn Ser Cys Lys Val Thr Val
355 360 365

Arg Lys Asn Tyr Arg Leu Leu Ser Cys Leu Val Gly Asp Ser Leu Leu
370 375 380

Glu Pro Phe Trp Pro Thr Gly Ser Gly Cys Ala Arg Gly Phe Leu Ser
385 390 395 400

Ser Met Asp Ala Ala Tyr Ala Ile Lys Leu Phe Ala Asn Pro Lys Asn
405 410 415

Ser Leu Leu Ala Thr Ile Ala Gln Arg Glu Ser Val Tyr Arg Leu Leu
420 425 430

Gly Gln Thr Thr Pro Glu Asn Leu Asn Arg Ala Phe Gly Ala Tyr Thr
435 440 445

Leu Asp Pro Ser Thr Arg Tyr Lys Asn Leu Asn Lys Ala Ser Val Gln
450 455 460

Ile Gly Gln Val Lys His Leu Leu Asp Thr Asp Asp
465 470 475

<210> 25
<211> 211
<212> PRT
<213> Ciona intestinalis

<400> 25

Asn Ile Val Tyr Tyr Lys Gly Glu Thr His Tyr Phe Val Met Thr Ala
1 5 10 15

Lys Lys His Ser Leu Val Ser Lys Gly Val Leu Lys Gln Asp Tyr Asp
20 25 30

Asn Thr Asn Glu Leu Leu Cys Tyr Asn Asn Ile Asp Gln Glu Glu Leu
35 40 45

Met Lys Tyr Ala Lys Gln Ala Ala Asp Phe Ser Thr Arg His Gln Leu
50 55 60

Pro His Leu Asp Phe Ala Ile Asn Gln Tyr Gly Gln Ser Asp Ile Ala
65 70 75 80

Leu Phe Asp Phe Thr Cys Met Tyr Ala Ala Glu Asn Ala Ala Leu Phe
85 90 95

Arg Glu Thr Tyr Arg Gln Lys Leu Leu Cys Cys Leu Val Gly Asp Ser
100 105 110

Leu Leu Glu Pro Phe Trp Pro Met Gly Thr Gly Cys Ala Arg Gly Phe
115 120 125

Leu Ala Ala Phe Asp Leu Val Trp Met Thr Lys Gln Leu Ala Leu Lys
130 135 140

Arg Lys Cys Ser Asn Tyr Asp Pro Asn Asp Asn Lys Val Glu Leu Ala
145 150 155 160

Val Leu Ala Glu Arg Glu Ser Ile Tyr Arg Val Leu His Gln Thr Thr
165 170 175

Pro Gln Asn Thr Met Lys Asn His Gln Asp Tyr Thr Ile Ala Pro Ser
180 185 190

Thr Arg Tyr Ala Asn Leu Asn Leu Lys Ala Val Thr Pro Ser Gln Val
195 200 205

Lys Pro Leu
210

<210> 26
<211> 252
<212> PRT
<213> Danio rerio

<400> 26

Phe Arg Arg Lys Glu Phe Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala
1 5 10 15

Asn Phe Ile Asn Arg Asn Thr Thr Ala Glu Ala Lys Val Glu Glu Ile
20 25 30

Ser Gly Val Ala Phe Ile Phe Asn Gln Lys Phe Phe Gln Asp Leu Arg
35 40 45

Glu Ala Thr Gly Ile Asp Leu Glu Asn Ile Val Tyr Tyr Lys Asp Asp
50 55 60

Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln Ser Leu Leu Glu Lys
65 70 75 80

Gly Val Ile Leu Asp Tyr Ala Asp Thr Glu Met Leu Leu Ser Arg Ala
85 90 95

Asn Val Asp Gln Lys Ala Leu Leu Ser Tyr Ala Arg Glu Ala Ala Asp
100 105 110

Phe Ser Thr Asn His Gln Leu Pro Lys Leu Asp Phe Ala Ile Asn His
115 120 125

Tyr Gly Gln Pro Asp Val Ala Met Phe Asp Phe Thr Cys Met Tyr Ala
130 135 140

Ser Glu Asn Ala Ala Leu Val Arg Gln Arg Asn Gly His Lys Leu Leu
145 150 155 160

Val Ala Leu Val Gly Asp Ser Leu Leu Glu Pro Phe Trp Pro Met Gly
165 170 175

Thr Gly Ile Ala Arg Gly Phe Leu Ala Ala Met Asp Ser Ala Trp Met
180 185 190

Val Arg Ser Trp Ala His Gly Ser Ser Pro Leu Glu Val Leu Ala Glu
195 200 205

Arg Glu Ser Ile Tyr Arg Leu Leu Pro Gln Thr Thr Pro Glu Asn Val
210 215 220

Ser Lys Asn Phe Ser Gln Tyr Ser Val Asp Pro Thr Thr Arg Tyr Pro
225 230 235 240

Asn Ile Ser Leu His Gln Val Arg Pro Asn Gln Val
245 250

<210> 27
<211> 154
<212> PRT
<213> Danio rerio

<400> 27

Ala Asp His Pro Val Ala Asp Tyr Asp Phe Asp Val Val Val Gly Ala
1 5 10 15

Asp Gly Arg Arg Asn Ser Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg
20 25 30

Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Thr Asn Arg Asn Thr
35 40 45

Thr Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe
50 55 60

Asn Gln Lys Phe Phe Gln Asp Leu Arg Gln Glu Thr Gly Ile Asp Leu
65 70 75 80

Glu Asn Ile Val Tyr Tyr Lys Asp Asn Thr His Tyr Phe Val Met Thr
85 90 95

Ala Lys Lys Gln Ser Leu Leu Asp Lys Gly Val Ile Ile His Asp Tyr
100 105 110

Ile Asp Thr Glu Ala Leu Leu Asn Ser Glu Asn Val Asn Gln Glu Ala
115 120 125

Leu Leu Val Tyr Ala Arg Glu Ala Ala Asp Tyr Ala Thr His Tyr Gln
130 135 140

Leu Pro Thr Leu Asp Tyr Ala Met Asn His
145 150

<210> 28
<211> 230
<212> PRT
<213> Gallus gallus

<400> 28

Leu Phe Asp Arg Phe Val Gln Ala Ser Thr Cys Lys Gly Thr Leu Lys
1 5 10 15

Ala Phe Gln Glu Leu Cys Asp Tyr Leu Glu Leu Lys Pro Lys Asp Tyr
20 25 30

Arg Ser Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala
35 40 45

Lys Ala Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr
50 55 60

Lys Lys Gly Lys Ala Cys Ala Asn Thr Lys Cys Leu Ile Ile Gly Ala
65 70 75 80

Gly Pro Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Phe Leu Gly Ala
85 90 95

Lys Val Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val
100 105 110

Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala
115 120 125

Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His Ile Ser
130 135 140

Ile Arg Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly
145 150 155 160

Ile Glu Ile His Val Asn Val Glu Phe Gln Gly Leu Val Tyr Pro Pro
165 170 175

Glu Asp Gln Glu Asn Glu Arg Ile Gly Trp Arg Ala Leu Val His Pro
180 185 190

Lys Thr His Pro Val Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly
195 200 205

Asp Gly Arg Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe Arg
210 215 220

Gly Lys Leu Ala Ile Ala
225 230

<210> 29
<211> 227
<212> PRT
<213> Gallus gallus

<400> 29

Leu Phe Glu His Phe Ile Arg Ala Arg Gln Cys Gln Glu Val Leu Ser
1 5 10 15

Cys Phe Ala Glu Leu Cys His Gln Leu Gly Leu Arg Gly Asn Gly Leu
20 25 30

Gln Leu Tyr His Ser Leu Lys Ala Ala Leu Asn Phe Trp Ser Ala Lys
35 40 45

Ala Leu Trp Ile Lys Leu Asp Lys Lys Ala Gly His Lys Asp Tyr Asp
50 55 60

Gln Gly Thr Ala Cys Ala Ser Thr Lys Cys Leu Val Val Gly Ala Gly
65 70 75 80

Pro Cys Gly Leu Arg Thr Ala Ile Glu Leu Ala Leu Leu Gly Ala Arg
85 90 95

Val Val Val Leu Glu Lys Arg Asp Ser Phe Ser Arg Asn Asn Val Leu
100 105 110

His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Ala Leu Gly Ala Lys
115 120 125

Lys Phe Tyr Gly Arg Phe Cys Thr Gly Thr Leu Asp His Ile Ser Ile
130 135 140

Arg Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Leu Leu Gly Val
145 150 155 160

Glu Val His Thr Lys Val Gln Phe Lys Gly Leu His Pro Pro Thr Gly
165 170 175

Lys Ala Ala Gly Gln Gly Gly Trp Arg Ala Val Leu Gln Pro Ser Ser
180 185 190

Ser Pro Leu Ser His Tyr Glu Phe Asp Val Leu Ile Ser Ala Gly Gly
195 200 205

Gly Lys Phe Val Pro Glu Asp Phe Lys Arg Lys Glu Met Arg Gly Lys
210 215 220

Leu Ala Ile
225

<210> 30
<211> 467
<212> PRT
<213> Rattus norvegicus

<400> 30

Gln Val Phe Glu Asn Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu
1 5 10 15

Gln Ala Phe Asn Ile Leu Thr Cys Leu Leu Asp Leu Asp Pro Leu Asp
20 25 30

His Arg Asn Phe Tyr Thr Gln Leu Lys Ser Lys Val Asn Thr Trp Lys
35 40 45

Ala Lys Ala Leu Trp His Lys Leu Asp Lys Arg Gly Ser His Lys Glu
50 55 60

Tyr Lys Arg Gly Lys Ala Cys Ser Asn Thr Lys Val Leu Ile Val Gly
65 70 75 80

Gly Gly Pro Cys Gly Leu Arg Thr Ala Ile Glu Leu Ala Tyr Leu Gly
85 90 95

Ala Lys Val Val Val Val Glu Lys Arg Asp Thr Phe Ser Arg Asn Asn
100 105 110

Val Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly
115 120 125

Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His Ile
130 135 140

Ser Ile Arg Gln Leu Gln Leu Ile Leu Phe Lys Val Ala Leu Met Leu
145 150 155 160

Gly Val Glu Ile His Val Asn Val Glu Phe Val Arg Val Arg Glu Pro
165 170 175

Pro Lys Asp Gly Trp Arg Ala Glu Phe Leu Pro Ala Asp His Ala Leu
180 185 190

Ser Asn Phe Glu Phe Asp Val Ile Ile Gly Ala Asp Gly His Arg Asn
195 200 205

Thr Leu Glu Phe Arg Arg Lys Glu Phe Arg Gly Lys Leu Ala Ile Ala
210 215 220

Ile Thr Ala Asn Phe Ile Asn Arg Asn Ser Thr Ala Glu Ala Lys Val
225 230 235 240

Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln Lys Phe Phe Gln

245 250 255

Asp Leu Lys Glu Glu Thr Gly Ile Asp Leu Glu Asn Ile Val Tyr Tyr
260 265 270

Lys Asp Ser Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln Ser Leu
275 280 285

Leu Asp Lys Gly Val Ile Leu Gln Asp Tyr Ile Asp Thr Glu Met Leu
290 295 300

Leu Cys Ala Glu Asn Val Asn Gln Asp Asn Leu Leu Ser Tyr Ala Arg
305 310 315 320

Glu Ala Ala Asp Phe Ala Thr Asn Tyr Gln Leu Pro Ser Leu Asp Phe
325 330 335

Ala Ile Asn His Asn Gly Gln Pro Asp Val Ala Met Phe Asp Phe Thr
340 345 350

Ser Met Tyr Ala Ser Glu Asn Ala Ala Leu Met Arg Glu Arg Gln Ala
355 360 365

His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu Leu Glu Pro Phe
370 375 380

Trp Pro Met Gly Thr Gly Cys Ala Arg Gly Phe Leu Ala Ala Phe Asp
385 390 395 400

Thr Ala Trp Met Val Lys Ser Trp Asp Gln Gly Thr Pro Pro Leu Glu
405 410 415

Val Leu Ala Glu Arg Glu Ser Leu Tyr Arg Leu Leu Pro Gln Thr Thr
420 425 430

Pro Glu Asn Ile Asn Lys Asn Phe Glu Gln Tyr Thr Leu Asp Pro Ala
435 440 445

Thr Arg Tyr Pro Asn Leu Asn Val His Cys Val Arg Pro His Gln Val
450 455 460

Ser Ala Leu
465

<210> 31
<211> 467

<212> PRT
<213> Rattus norvegicus

<400> 31

Phe Glu Thr Phe Val Gln Ala Gln Leu Cys Gln Asp Val Leu Ser Ser
1 5 10 15

Phe Gln Gly Leu Cys Arg Ala Leu Gly Val Glu Ser Gly Gly Gly Leu
20 25 30

Pro Gln Tyr His Lys Ile Lys Ala Gln Leu Asn Tyr Trp Ser Ala Lys
35 40 45

Ser Leu Trp Ala Lys Leu Asp Lys Arg Ala Ser Gln Pro Ala Tyr Gln
50 55 60

Gln Gly Gln Ala Cys Thr Asn Thr Lys Val Leu Val Val Gly Ala Gly
65 70 75 80

Pro Cys Gly Leu Arg Ala Ala Val Glu Leu Ala Leu Leu Gly Ala Arg
85 90 95

Val Val Leu Val Glu Lys Arg Thr Lys Phe Ser Arg His Asn Val Leu
100 105 110

His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Ala Leu Gly Ala Lys
115 120 125

Lys Phe Tyr Gly Arg Phe Cys Thr Gly Thr Leu Asp His Ile Ser Ile
130 135 140

Arg Gln Leu Gln Leu Leu Leu Leu Lys Val Ala Leu Leu Leu Gly Val
145 150 155 160

Glu Ile His Trp Gly Phe Thr Phe Thr Gly Leu Gln Pro Pro Pro Lys
165 170 175

Lys Gly Gly Ser Gly Trp Arg Ala Arg Ile Gln Pro Ser Pro Pro Ala
180 185 190

Gln Leu Ala Ser Tyr Glu Phe Asp Val Leu Ile Ser Ala Gly Gly Gly
195 200 205

Lys Phe Val Leu Gly Phe Thr Ile Arg Glu Met Arg Gly Lys Leu Ala
210 215 220

Ile Gly Ile Thr Ala Asn Phe Val Asn Gly Arg Thr Val Glu Glu Thr
225 230 235 240

Gln Val Pro Glu Ile Ser Gly Val Ala Arg Ile Tyr Asn Gln Lys Phe
245 250 255

Phe Gln Ser Leu Leu Lys Ala Thr Gly Ile Asp Leu Glu Asn Ile Val
260 265 270

Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln
275 280 285

Cys Leu Leu Arg Leu Gly Val Leu Arg Gln Asp Leu Pro Glu Thr Asp
290 295 300

Gln Leu Leu Gly Lys Ala Asn Val Val Pro Glu Ala Leu Gln Gln Phe
305 310 315 320

Ala Arg Ala Ala Ala Asp Phe Ala Thr Gln Gly Lys Leu Gly Lys Leu
325 330 335

Glu Phe Ala Gln Asp Ala Arg Gly Arg Pro Asp Val Ala Ala Phe Asp
340 345 350

Phe Thr Ser Met Met Arg Ser Glu Ser Ser Ala Arg Ile Gln Glu Lys
355 360 365

His Gly Ala Arg Leu Leu Leu Gly Leu Val Gly Asp Cys Leu Val Glu
370 375 380

Pro Phe Trp Pro Leu Gly Thr Gly Val Ala Arg Gly Phe Leu Ala Ala
385 390 395 400

Phe Asp Ala Ala Trp Met Val Lys Arg Trp Ala Glu Gly Thr Gly Pro
405 410 415

Leu Glu Leu Leu Ala Glu Arg Glu Ser Leu Tyr Gln Leu Leu Ser Gln
420 425 430

Thr Ser Pro Glu Asn Met His Arg Asn Val Ala Gln Tyr Gly Leu Asp
435 440 445

Pro Ala Thr Arg Tyr Pro Asn Leu Asn Leu Arg Ala Val Thr Pro Asn
450 455 460

Gln Val Arg

465

<210> 32
<211> 468
<212> PRT
<213> Rattus norvegicus

<400> 32

Leu Phe Asp Arg Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Arg
1 5 10 15

Ala Phe Gln Glu Leu Cys Asp His Leu Glu Leu Lys Pro Lys Asp Tyr
20 25 30

Arg Ser Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala
35 40 45

Lys Ala Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr
50 55 60

Lys Lys Gly Lys Ala Cys Thr Asn Thr Lys Val Leu Ile Ile Gly Ala
65 70 75 80

Gly Pro Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Leu Leu Gly Ala
85 90 95

Lys Val Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val
100 105 110

Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala
115 120 125

Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ala Ile Asp His Ile Ser
130 135 140

Ile Arg Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly
145 150 155 160

Ile Glu Ile His Val Asn Val Glu Phe Gln Gly Leu Val Gln Pro Pro
165 170 175

Glu Asp Gly Ile Gly Trp Arg Ala Leu Val His Pro Lys Thr His Pro
180 185 190

Val Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly Asp Gly Arg Arg
195 200 205

Asn Thr Leu Glu Phe Arg Arg Lys Glu Phe Arg Gly Lys Leu Ala Ile
210 215 220

Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Thr Thr Ala Glu Ala Lys
225 230 235 240

Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln Lys Phe Phe
245 250 255

Gln Glu Leu Arg Glu Ala Thr Gly Gly Ile Asp Leu Glu Asn Ile Val
260 265 270

Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln
275 280 285

Ser Leu Leu Asp Lys Gly Val Ile Leu Gln Asp Tyr Thr Asp Thr Glu
290 295 300

Leu Leu Leu Ser Arg Glu Asn Val Asp Gln Glu Ala Leu Leu Asn Tyr
305 310 315 320

Ala Arg Glu Ala Ala Asp Phe Ser Thr Gln Gln Gln Leu Pro Ser Leu
325 330 335

Asp Phe Ala Ile Asn His Tyr Gly Gln Pro Asp Val Ala Met Phe Asp
340 345 350

Phe Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg Glu Gln
355 360 365

Asn Gly His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu Leu Glu
370 375 380

Pro Phe Trp Pro Met Gly Thr Gly Ile Ala Arg Gly Phe Leu Ala Ala
385 390 395 400

Met Asp Ser Ala Trp Met Val Arg Ser Trp Ser Leu Gly Thr Ser Pro
405 410 415

Leu Glu Val Leu Ala Glu Arg Arg Glu Ser Ile Tyr Arg Leu Leu Pro
420 425 430

Gln Thr Thr Pro Glu Asn Val Ser Lys Asn Phe Ser Gln Tyr Ser Ile
435 440 445

Asp Pro Val Thr Arg Tyr Pro Asn Ile Asn Ile Asn Phe Leu Arg Pro
450 455 460

Ser Gln Val Arg
465

<210> 33
<211> 428
<212> PRT
<213> Bos taurus

<400> 33

Leu Phe Asp Arg Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Lys
1 5 10 15

Ala Phe Gln Glu Leu Cys Asp His Leu Glu Leu Lys Pro Lys Asp His
20 25 30

Arg Ser Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala
35 40 45

Lys Ala Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr
50 55 60

Lys Lys Gly Lys Val Cys Thr Asn Thr Lys Val Leu Ile Ile Gly Ala
65 70 75 80

Gly Pro Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Leu Leu Gly Ala
85 90 95

Lys Val Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val
100 105 110

Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala
115 120 125

Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ala Ile Asp His Ile Ser
130 135 140

Arg Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly Ile
145 150 155 160

Glu Ile His Val Asn Val Glu Phe Arg Gly Leu Val Glu Pro Pro Glu
165 170 175

Asp Gly Ile Gly Trp Arg Ala Leu Val His Pro Lys Thr His Pro Val
180 185 190

Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly Asp Gly Arg Arg Asn
195 200 205

Thr Leu Glu Phe Arg Arg Lys Glu Phe Arg Gly Lys Leu Ala Ile Ala
210 215 220

Ile Thr Ala Asn Phe Ile Asn Arg Asn Thr Thr Ala Glu Ala Lys Val
225 230 235 240

Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln Lys Phe Phe Gln
245 250 255

Glu Leu Arg Glu Ala Thr Gly Gly Ile Asp Leu Glu Asn Ile Val Tyr
260 265 270

Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln Ser
275 280 285

Leu Leu Asp Lys Gly Val Ile Leu Gln Asp Tyr Ala Asp Thr Glu Leu
290 295 300

Leu Leu Ser Arg Glu Asn Val Asp Gln Glu Ala Leu Leu Ser Tyr Ala
305 310 315 320

Arg Glu Ala Ala Asp Phe Ser Thr Gln Gln Gln Leu Pro Ser Leu Asp
325 330 335

Phe Ala Ile Asn His Tyr Gly Gln Pro Asp Val Ala Met Phe Asp Phe
340 345 350

Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg Glu His Asn
355 360 365

Gly His Gln Leu Ala Trp Trp Leu Trp Val Gly Gly Asp Ser Leu Arg
370 375 380

Glu Ser Ile Tyr Arg Leu Leu Pro Gln Thr Thr Pro Glu Asn Val Ser
385 390 395 400

Lys Asn Phe Ser Gln Tyr Ser Ile Asp Pro Val Thr Arg Tyr Pro Asn
405 410 415

Val Asn Val Asn Phe Leu Arg Pro Ser Gln Val Arg
420 425

<210> 34
<211> 177
<212> PRT
<213> Bos taurus

<400> 34

Ile Thr Ala Asn Phe Val Asn Gly Arg Thr Val Glu Glu Thr Gln Val
1 5 10 15

Pro Glu Ile Ser Gly Val Ala Arg Ile Tyr Asn Gln Ser Phe Phe Gln
20 25 30

Ser Leu Leu Lys Ala Thr Gly Ile Asp Leu Glu Asn Ile Val Tyr Tyr
35 40 45

Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln Cys Leu
50 55 60

Leu Arg Leu Gly Val Leu His Lys Asp Trp Pro Asp Thr Glu Arg Leu
65 70 75 80

Leu Gly Ser Ala Asn Val Val Pro Glu Ala Leu Gln Arg Phe Ala Arg
85 90 95

Ala Ala Ala Asp Phe Ala Thr His Gly Lys Leu Gly Lys Leu Glu Phe
100 105 110

Ala Arg Asp Ala His Gly Arg Pro Asp Val Ser Ala Phe Asp Phe Thr
115 120 125

Ser Met Met Arg Ala Glu Ser Ser Ala Arg Val Gln Glu Arg His Gly
130 135 140

Thr Arg Leu Leu Leu Gly Leu Val Gly Asp Cys Leu Val Glu Pro Phe
145 150 155 160

Trp Pro Leu Gly Thr Gly Val Ala Arg Gly Phe Leu Ala Ala Phe Asp
165 170 175

Ala

<210> 35
<211> 169
<212> PRT
<213> Sus scrofa

<400> 35

Ala Lys Val Val Val Val Glu Lys Arg Asp Thr Phe Ser Arg Asn Asn
1 5 10 15

Val Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly
20 25 30

Ala Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ser Ile Asp His Ile
35 40 45

Ser Ile Arg Gln Leu Gln Leu Ile Leu Phe Lys Val Ala Leu Leu Leu
50 55 60

Gly Val Glu Ile His Val Asn Val Glu Phe Val Lys Val Leu Glu Pro
65 70 75 80

Pro Glu Asp Gln Glu Asn Gln Lys Ile Gly Trp Arg Ala Glu Phe Leu
85 90 95

Pro Ala Asp His Ser Leu Ser Glu Phe Glu Phe Asp Val Ile Ile Gly
100 105 110

Ala Asp Gly Arg Arg Asn Thr Leu Glu Gly Phe Arg Arg Lys Glu Phe
115 120 125

Arg Gly Lys Leu Ala Ile Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn
130 135 140

Ser Thr Ala Glu Ala Lys Val Glu Glu Ile Ser Gly Val Ala Phe Ile
145 150 155 160

Phe Asn Gln Lys Phe Phe Gln Asp Leu
165

<210> 36
<211> 468
<212> PRT
<213> Pan troglodytes

<220>
<221> MISC_FEATURE
<222> (298)..(298)
<223> Xaa is any amino acid

<400> 36

Leu Phe Asp Arg Phe Val Gln Ala Thr Thr Cys Lys Gly Thr Leu Lys
1 5 10 15

Ala Phe Gln Glu Leu Cys Asp His Leu Glu Leu Lys Pro Lys Asp Tyr
20 25 30

Arg Ser Phe Tyr His Lys Leu Lys Ser Lys Leu Asn Tyr Trp Lys Ala
35 40 45

Lys Ala Leu Trp Ala Lys Leu Asp Lys Arg Gly Ser His Lys Asp Tyr
50 55 60

Lys Lys Gly Lys Ala Cys Ala Asn Thr Lys Val Leu Ile Ile Gly Ala
65 70 75 80

Gly Pro Cys Gly Leu Arg Thr Ala Ile Asp Leu Ser Leu Leu Gly Ala
85 90 95

Lys Val Val Val Ile Glu Lys Arg Asp Ala Phe Ser Arg Asn Asn Val
100 105 110

Leu His Leu Trp Pro Phe Thr Ile His Asp Leu Arg Gly Leu Gly Ala
115 120 125

Lys Lys Phe Tyr Gly Lys Phe Cys Ala Gly Ala Ile Asp His Ile Ser
130 135 140

Ile Arg Gln Leu Gln Leu Ile Leu Leu Lys Val Ala Leu Ile Leu Gly
145 150 155 160

Ile Glu Ile His Val Asn Val Glu Phe Gln Gly Leu Ile Gln Pro Pro
165 170 175

Glu Asp Gly Ile Gly Trp Arg Ala Leu Val His Pro Lys Thr His Pro
180 185 190

Val Ser Glu Tyr Glu Phe Glu Val Ile Ile Gly Gly Asp Gly Arg Arg
195 200 205

Asn Thr Leu Glu Phe Arg Arg Lys Glu Phe Arg Gly Lys Leu Ala Ile
210 215 220

Ala Ile Thr Ala Asn Phe Ile Asn Arg Asn Thr Thr Ala Glu Ala Lys
225 230 235 240

Val Glu Glu Ile Ser Gly Val Ala Phe Ile Phe Asn Gln Lys Phe Phe
245 250 255

Gln Glu Leu Arg Glu Ala Thr Gly Gly Ile Asp Leu Glu Asn Ile Val
260 265 270

Tyr Tyr Lys Asp Asp Thr His Tyr Phe Val Met Thr Ala Lys Lys Gln
275 280 285

Ser Leu Leu Asp Lys Gly Val Ile Leu Xaa Asp Tyr Ala Asp Thr Glu
290 295 300

Leu Leu Leu Ser Arg Glu Asn Val Asp Gln Glu Ala Leu Leu Ser Tyr
305 310 315 320

Ala Arg Glu Ala Ala Asp Phe Ser Thr Gln Gln Gln Leu Pro Ser Leu
325 330 335

Asp Phe Ala Ile Asn His Tyr Gly Gln Pro Asp Val Ala Met Phe Asp
340 345 350

Phe Thr Cys Met Tyr Ala Ser Glu Asn Ala Ala Leu Val Arg Glu Gln
355 360 365

Asn Gly His Gln Leu Leu Val Ala Leu Val Gly Asp Ser Leu Leu Glu
370 375 380

Pro Phe Trp Pro Met Gly Thr Gly Ile Ala Arg Gly Phe Leu Ala Ala
385 390 395 400

Met Asp Ser Ala Trp Met Val Arg Ser Trp Ser Leu Gly Thr Ser Pro
405 410 415

Leu Glu Val Leu Ala Glu Arg Arg Glu Ser Ile Tyr Arg Leu Leu Pro
420 425 430

Gln Thr Thr Pro Glu Asn Val Ser Lys Asn Phe Ser Gln Tyr Ser Ile
435 440 445

Asp Pro Val Thr Arg Tyr Pro Asn Ile Asn Val Asn Phe Leu Arg Pro
450 455 460

Ser Gln Val Arg
465

<210> 37
<211> 18
<212> DNA
<213> Artificial sequence

<220>
<223> Amplification primer

<400> 37
ggagcagggc cctgtgga 18

<210> 38
<211> 18
<212> DNA
<213> Artificial sequence

<220>
<223> Amplification primer

<400> 38
tgggcatggc cctgttgg 18

<210> 39
<211> 7
<212> PRT
<213> Drosophila

<400> 39

Gly Ala Gly Pro Cys Gly Leu
1 5

<210> 40
<211> 7
<212> PRT
<213> Artificial sequence

<220>
<223> FAD-binding domain mutant

<400> 40

Trp Ala Trp Pro Cys Trp Leu
1 5